EFFECTS OF SEMANTIC AND SEGMENTAL SIMILARITY ON THE PRODUCTION AND LEARNING OF SPOKEN AND WRITTEN WORDS

by
Bonnie Lynn Breining

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
October, 2016

ABSTRACT

This dissertation investigates mechanisms of word production and learning, focusing on how semantic and segmental similarity affect the production and learning of words. Incremental learning models of spoken word production propose that learning occurs each time a familiar word is produced. In recent work, I argued that this type of model applies across stages and modalities of word production. Here, I extend the model to the learning of new words, incorporating insights from the literature regarding learning difficulty to make predictions about training effects and long-term learning outcomes. The extended incremental learning model (e-ILM) framework I propose makes testable predictions about the effects of training new words in semantic or segmentally related blocks vs. unrelated contexts. According to these predictions, semantic blocking is a desirable difficulty that causes interference during training but improves long-term learning by enhancing the distinctiveness of the learned representations. In contrast, segmental blocking is predicted to negatively affect both training and long-term learning because it reduces distinctiveness.

Two studies tested these predictions. In the first, neurotypical adults learned names for novel items in semantic, segmental, and unrelated blocks in separate written and spoken experiments. Findings were consistent with the e-ILM predictions about training: although training in both types of related vs. unrelated blocks produced interference, semantic blocking increased distinctiveness while segmental blocking reduced distinctiveness. The predictions regarding long-term learning were less clearly supported because blocking did not consistently affect retention.

In the second study, individuals with acquired dysgraphia relearned word spellings in semantic, segmental, and unrelated blocks. Group results did not consistently support the e-ILM predictions about training. However, there was evidence that both semantic and segmental blocking acted as desirable difficulties that improved long-term retention to the degree to which they increased training difficulty, suggesting that the general learning principle of desirable difficulty applies in this situation.

Overall, the proposed e-ILM account accurately describes the underlying mechanisms that produce effects of blocking by semantic or segmental similarity during the (re)learning of words. The work raises potentially important questions regarding differences between new learning and relearning in neurotypical and brain damaged individuals that motivate further research.

Committee Members

Brenda Rapp (advisor)

Argye Hillis

Michael McCloskey

Nazbanou Nozari

Colin Wilson

ACKNOWLEDGMENTS

I am immensely grateful to the many people helped to make this dissertation a reality. Below, I wish to thank a few of the most salient ones.

To my advisor, Brenda Rapp: Thank you for your patience. Thank you for helping me develop precision and perseverance, which positively affect my life in science and beyond. You have taught me to speak, write, and, most of all, think more carefully and clearly. I appreciate all your encouragement to continue my endeavors until they are not just done, but done well.

To my other committee members, Argye Hillis, Bonnie Nozari, Colin Wilson, and Mike McCloskey: Thank you for your time and for productively engaging with my work, not just on this dissertation but throughout my graduate school education.

To the members of the Cognitive Science Department: Thank you for your intellectual generosity and for ensuring that I have not just depth but also breadth in my education.

To those who helped complete the research presented here: It would have been impossible without you. To DDR, DWS, ESG, GHN, KSR2, REN, and SMY: it was a joy to get to know each of you as we worked together. Thank you for persevering through the many boring sessions of writing the same words over and over. To Denise McCall, Lisa Thornburg, Jes Porro, and the many students working at SCALE during my study: Thank you for letting me hijack your idea for a spelling class, doing all the behind-the-scenes work to make my study and its complex scheduling demands possible, and still acting like I was doing you a favor. To the many JHU students who participated in my experiments: A major part of success is showing up, and your showing up directly lead to the success of my project. To the numerous research assistants who worked with me, including Alex Surget, Alexandra Gordon, Christie Okoye, Elena Gresick, Haley Wendt, Helena Chung, Julie Paone, Malena Silva, Michelle Chiu, Namir Khalasawi, Sue Junn, and Vyshnavi Anandan: Your contributions made this project manageable.

To my fellow graduate students and lab mates past and present, particularly members of the CogNeuro Lab and my cohort: Thank you for being good colleagues and even better friends. Special shout-outs to those of you whose support extends well beyond academic life, including (but not limited to) Chris Kirov, David Rothlein, Deepti Ramadoss, Emma Gregory, Eleanor Chodroff, Erin Zaroukian, Jen Shea, Kristen Johannes, Manny Vindiola, and Teresa Schubert.

To my parents, Linda and John Breining: Thank you for teaching me the value of a good education and making sure I got one. Thank you for believing I could do whatever I set my mind to, and for pushing me to do my best. Thanks for always acting interested in what I'm doing, even when I'm not that interested in talking about it.

To my husband, Shawn Cherian: Thanks for reminding me about all the awesome things outside of work and making me participate in them. You've only known me as a graduate student, and I appreciate that you still want to hang out with me after all of this. Thank you for always knowing I could do it, and for supporting me in all my efforts so far and in those still to come.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

OVERVIEW

In this dissertation, I consider the effects of producing and learning words in the context of other words that are semantically or segmentally related.

In Chapter 1, I discuss background work arguing that incremental learning models previously proposed to account for semantic interference effects in spoken production can be extended across stages and modalities of production. This type of model, which proposes that learning occurs each time a word is produced, can be applied beyond lexical selection to segmental encoding in order to account for the interference observed when words are produced in the context of other items that share many of its segments. Furthermore, such models apply to both spoken and written production, explaining consistent interference effects observed in both modalities.

In the rest of the dissertation, I consider the implications that these models of production have for the learning of new items and the relearning of previously known items. In Chapter 2, I review literature drawn from the fields of education, rehabilitation, and second language learning regarding the role of difficulty in learning. Across many situations, increasing retrieval difficulty during acquisition of knowledge and skills leads to positive long-term learning outcomes. However, the mechanisms that lead to these positive effects have not been elucidated. In Chapter 3, I propose a framework for understanding the underlying mechanisms that lead to observable effects of increasing difficulty in word learning by training words in related vs. unrelated blocks. I suggest that an incremental learning model predicts different effects of semantic as opposed to segmental similarity. According to this model, learning groups of items that are semantically similar will present a desirable difficulty: semantic blocking may initially cause interference but have beneficial effects on long term learning by enhancing the distinctiveness of the items. On the other hand, I predict that learning groups of items that are segmentally similar will have negative effects both during training and when long term leaning is assessed because this type of learning context may reduce distinctiveness of the items.

I then present two studies that test these predictions. In the first study, reported in Chapter 4, neurologically healthy adults learned new names for novel items in blocks of semantically related, segmentally related, and unrelated items in separate written and spoken experiments. In this study, I assessed the effects that blocking has on the trajectory and outcomes of learning, examining both production speed and accuracy. Furthermore, I examined the e-ILM predictions regarding changes to the distinctiveness of representations as a result of blocking through the results of verification probe tasks that compare distinctive and shared features of items trained in different contexts.

In the second study, presented in Chapter 5, individuals with acquired dysgraphia relearned previously known spellings for common items in blocks of semantically related, segmentally related, and unrelated items. In this study, I also examined how the effects of blocking are impacted by the various cognitive deficits with which participants present. I assessed the effects of blocking on the trajectory and outcomes of relearning, examining production accuracy.

Finally, Chapter 6 presents a summary of the dissertation. Overall, the work presented here expands our understanding of how the word production system works and how new words are learned. The results provide insights that can be applied to enhance learning and relearning of words.

CHAPTER 1: MECHANISMS OF THE LANGUAGE PROCESSING SYSTEM

On Scope

Cognitive Architecture of Production

One of the goals of this dissertation is to better understand the mechanisms underlying the word production system, both as they apply to the production of known words and the learning of new words. With this goal in mind, it is helpful to first consider the architecture of the single word production system, presented in Figure 1. (For review of the system in line with the description I present below, see Buchwald & Rapp, 2009; Ellis & Young, 1988; Goldrick & Rapp, 2007; Nickels, 2001; Tainturier & Rapp, 2001.)

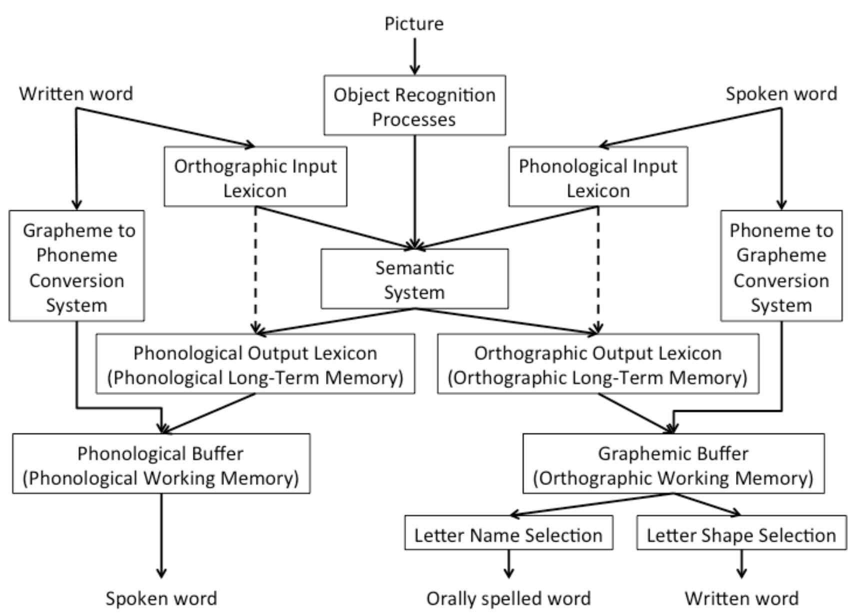


Figure 1. A cognitive architecture of the spoken and written word production system.

Written production. I begin by describing written production. In order to spell the name of a picture, the visual input from the picture must first undergo object recognition processes. Once the picture is recognized, the stored representation of its meaning in the semantic system is activated. This in turn activates the long-term orthographic memory representation of the word's spelling in the orthographic output lexicon, which stores the written forms of known words. Because stored word form representations are contacted, picture naming is considered a lexical process. The graphemes that make up the spelling are then processed by orthographic working memory, a primary component of which is the graphemic buffer. The graphemic buffer maintains activation of the abstract representations of graphemes throughout production of the word being spelled. These abstract representations include information about the identity and serial position of the graphemes. Given this information, the letter names for the graphemes can be selected, and the motor programs used to pronounce those letter names activated so that the word can be spelled aloud. Alternatively, the letter shapes for the graphemes can be selected, followed by activation of motor programs, so that the word can be written.

In order to spell a word given its spoken name (i.e., spelling to dictation), two routes of production can be used. The first is the lexical route. After auditory processing of the heard stimulus, the stored phonological form of the word (long-term phonological memory representation) is activated in the phonological input lexicon. Activation is then sent to the word meaning in the semantic system. From this point on, the process follows the same steps as written picture naming. Activation is sent from the semantic system to the long-term orthographic memory representation of the word, which is then processed by orthographic working memory. From there, letter name or letter shape selection processes are used to transform representations of the graphemes for oral or written spelling. Note that some have also proposed a non-semantically mediated lexical route that directly links the phonological input lexicon and the orthographic output lexicon (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; although see Hillis & Caramazza, 1991), which is depicted by a dashed line in Figure 1. Alternatively, the sublexical route can be used. After auditory processing, this route directly converts the phonemes of the auditory stimulus to graphemes, based on plausible, regular mappings of phonology to orthography. The graphemes that are generated via this phoneme-to-grapheme conversion process are processed in orthographic working memory, and production of the spelling after letter name selection or letter shape selection proceeds as in the lexical route. This sublexical route is useful because it allows for the spelling of unknown words that do not have stored meanings or forms.

Spoken production. The spoken production system is analogous to the written production system. In order to say the name of a picture, the same object recognition processes and activation of the stored representation of the recognized picture's meaning in the semantic system must occur as in spelling the name of a picture. Differences arise after this point because the output modality is spoken as opposed to written. Producing spoken output requires accessing the long-term phonological memory representation of the spoken word form in the phonological output lexicon instead of the long-term orthographic memory representation. The phonemes that make up the representation of the spoken word form are then processed by phonological working memory (instead of orthographic working memory) throughout production of the word, which occurs as a result of articulatory planning and execution of motor programs.

The process of saying a word given its written name (i.e., reading aloud) is analogous to the process of spelling a word given its spoken name (i.e., spelling to dictation). Again, there are two pathways of production, the lexical and sublexical routes. Both routes begin with visual processing to recognize the letters that make up the written word form. When the lexical route is utilized, the long-term orthographic memory representation is activated in the orthographic input lexicon. From this point forward, the same processing is used as in spoken picture naming, with activation of the semantic representation followed by activation of the long-term phonological memory representation in the phonological output lexicon (or use of the direct lexical route that bypasses the semantic system) and subsequent activation of the constituent phonemes in phonological working memory. When the sublexical route is used, the graphemes of the written word are directly converted to phonemes based on stored regular mappings of written forms to sound. These phonemes are held in phonological working memory, the same phonological buffer used by the lexical route. As in spoken picture naming, the phonemes held in phonological working memory as a result of processing in both the lexical and sublexical routes are utilized by articulatory planning processes that feed motor plans for pronunciation of the word.

Investigating lexical production. This dissertation focuses on the lexical production system, which has two primary stages: lexical selection and segmental encoding. Lexical selection is the process of identifying a specific lexical node to convey the intended meaning. Segmental encoding is the process of selecting the segments that make up the selected word, which consist of phonemes in spoken production and graphemes in written production. In the model of production presented in Figure 1 and described above, these stages are collapsed as processing at the level of the lexicon: they are involved in retrieving the long-term memory representations of words when given semantic information, and the information they generate is then processed by working memory systems.

The work reported in this dissertation investigates production via picture naming tasks. In the background experiments presented later in this chapter, I examined the effects of naming pictures in contexts where all words that are to be produced in a block have high semantic or segmental similarity as opposed to unrelated contexts where the words to be produced in a block have low semantic and segmental similarity. In the studies described in later chapters, participants learned or relearned the names of pictures in contexts where other items to be learned are semantically similar, segmentally similar, or unrelated. Picture naming is an appropriate choice for investigating the lexical production system because it is a lexical process that requires semantic processing and access to (or creation of) stored word forms; it cannot be accomplished through sublexical processing alone. Using picture naming tasks allows us to specifically investigate the lexical production system.

A model of lexical selection and segmental encoding. Based on the lexical focus of the dissertation, I present a focused model of the processes of interest. This model is in line with previously proposed models of lexical processing that have been implemented in computer simulations (e.g., Dell, Nozari, & Oppenheim, 2014; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Oppenheim, Dell, & Schwartz, 2010; Rapp &

Goldrick, 2000). Figure 2 depicts an example of the model, illustrating retrieval of a single word's long-term memory representation for production. Although written segments are depicted here, the models of spoken and written production operate analogously. When a picture of an object is shown (e.g., a cat), its semantic representation is activated. In this simplified model, this is depicted as activation of features such as "furry", "pet", and "meows". The semantic level corresponds to the semantic system of the cognitive architecture of word production presented above. The activated semantic features in turn send activation to the lexical level representation *cat*. Other lexical representations that share those features would also receive activation (e.g., "furry" and "pet" might also activate *dog*, although this is not depicted in Figure 2). The process of choosing amongst multiple active lexical representations is called lexical selection. The lexical level representations are part of the long-term memory representations of phonological and orthographic forms stored in the orthographic and phonological output lexicons in the cognitive architecture of word production presented above.[1] I refer to the stored word representations at the lexical level as lexical nodes. The

[1] In past research there has been debate about the format of representations at the lexical level. While there is agreement that there is at least one intervening level between conceptual-semantic representations and segmental representations, different proposals have been made regarding the nature of such lexical representations (e.g., Caramazza, 1997 for review of models with and without lemmas). These proposals differ in terms of whether the lexical representations are amodal or modality-specific and whether or not they contain syntactic information. Some proposals suggest there are lemmas, which consist of nodes for each lexical items that incorporate syntactic information. Accessing the segments of the word necessarily means accessing syntactic information about the word. Such lemmas are amodal, meaning that there is a single lemma that is contacted regardless of the modality of production: the same node is used in retrieval of both phonological and orthographic information. On the other hand, other proposals suggest that lexical representations are not lemmas but rather modality-specific lexemes. In a model with lexemes, separate nodes are used to retrieve phonological word forms and orthographic word forms (i.e., separate nodes mediate the connections between semantic representations and phonemes vs. graphemes). These lexemes themselves do not contain syntactic information (although they are likely connected to other nodes that represent syntactic features), which means that the segments of the word can be accessed without accessing syntax. Some models also incorporate both lemmas and lexemes, proposing that there are two levels of lexical representation: first the amodal lemmas are contacted, then the modality-specific lexemes.

The work presented in this dissertation does not depend on whether there are amodal lemmas, modality-specific lexemes, or both. The discussion here considers the representations at the lexical level to active lexical nodes next send activation to constituent segments through the process of segmental encoding. In this example, this leads to selection of the segments C, A, and T at the segmental level. The segmental level representations, which consist of graphemes for written production and phonemes for spoken production, are also part of the long-term memory representations of word forms stored in the lexicons. Working memory, which includes the graphemic and phonological buffers, processes the outputs of segmental encoding as production continues. The work in this dissertation will focus on the mechanisms underlying the processes of lexical selection and segmental encoding depicted here.

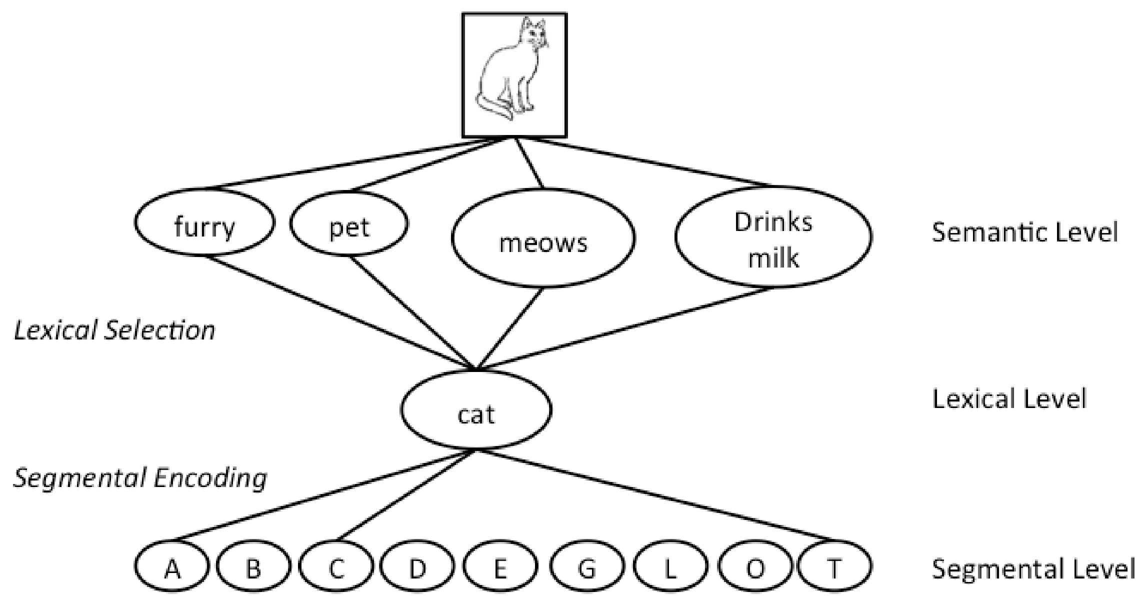


Figure 2. Architecture of lexical processes of word production.

---

be modality-specific phonological and orthographic lexemes. Lexical selection connects semantic features to modality-specific phonological and orthographic lexemes. Segmental encoding connects these modality-specific lexemes to the corresponding segments (phonemes or graphemes). This description of lexical representations as modality-specific lexemes is a simplifying assumption: The same mechanisms I discuss can be applied to the other types of representations as well. If the representations are abstract lemmas, lexical selection connects semantic features to lemmas, and segmental encoding connects the same lemmas to both modality-specific phonemes and graphemes. Task demands determine whether graphemes or phonemes are used. If both lemmas and lexemes are present, lexical selection connects semantic features to lemmas. Abstract lemmas are in turn connected to modality-specific lexemes. Segmental encoding connects lexemes to segments.

Spoken and Written Production: Functional Autonomy, Yet Similar General Principles

As is apparent from the discussion of production systems above, I will investigate both spoken and written production in this dissertation. Although written language depends on spoken language during development, numerous findings support the autonomy of written word production in the adult expert system (Rapp & Fischer-Baum, 2014). While one productive line of psycholinguistic research with neurologically healthy adults indicates that phonological information is automatically activated when orthographic information is activated (Alario, De Cara, & Ziegler, 2007; Humphreys, Evett, & Taylor, 1982; Rastle & Brysbaert, 2006; Zhang & Damian, 2010), phonological mediation between meaning and orthographic lexical retrieval is not obligatory. For example, Bonin, Fayol, and Peereman (1998) found that written picture naming was facilitated by the presentation of orthographically related pseudoword primes, regardless of whether the orthographically matched primes were pseudohomophones of the target. That is, there was no additional speed advantage in naming the target (e.g., chat /ʃa /) when it was preceded by a phonologically homophonous and orthographically similar prime (e.g., chax /ʃa /) compared to a non-homophone prime that was of comparable orthographical similarity (e.g., chap /ʃap /). In effect, facilitating retrieval of the phonological form did not facilitate retrieval of the orthographic form, in contrast to would be expected if phonological mediation were mandatory for written production. Such results suggest that individuals can directly retrieve word spellings from semantic representations without first retrieving the words' phonological forms.

The case for orthographic autonomy is further bolstered by neuropsychological evidence. A number of cases have been reported in which individuals impaired in lexical selection in spoken production exhibit relatively more intact lexical selection in the written modality (Ellis & Young, 1988; Miceli & Capasso, 1997; Rapp, Benzing, & Caramazza, 1997; Tainturier, Moreaud, David, Leek, & Pellat, 2001). One example is case R.G.B. (Caramazza & Hillis, 1990), who made semantic errors in spoken picture naming (e.g., picture of goose→"turkey") but not in written picture naming. If written lexical access in adulthood were necessarily mediated by spoken lexical access, these results would not be expected. There is also evidence that morphological processing is independent in the two modalities. For example, Rapp, Fischer-Baum, and Miozzo (2015) reported five individuals with aphasia who demonstrated a double dissociation whereby some individuals were more impaired in production of affixes in the spoken modality, whereas others were more impaired in the written modality.

Despite this functional autonomy, similar organizational principles are observed across spoken and written production (e.g., Bonin & Fayol, 2000; Shen, Damian, & Stadthagen-Gonzalez, 2013). For instance, grammatical category seems to play a similar role in the organization of lexical entries for both spoken and written production (e.g., Rapp & Caramazza, 1997). Furthermore, activation of non-target words is driven by consistent factors during both lexical phonological and lexical orthographic processing, including lexical frequency, grammatical category, target length, position-specific form overlap, and initial segment identity (Goldrick, Folk, & Rapp, 2010). Similarly, the same factors, including image variability, image agreement and age of acquisition, drive spoken and written picture naming latencies (Bonin, Chalard, Méot, & Fayol, 2002).

Therefore, it is reasonable to expect similarity in other principles governing written and spoken production.

In summary, spoken and written production follow a similar set of general cognitive principles, but not because written production relies on spoken production in the adult system. This means that one cannot assume that principles observed in spoken production necessarily generalize to written production. Written production itself must be investigated in addition to spoken production. Observing comparable effects in both modalities would suggest that core cognitive principles are domain-general, even though they operate on domain-specific representations. One aim of this dissertation is to investigate the extent to which activation and selection principles are shared across modalities. If there are in fact similar organizational principles, extending the work beyond spoken production to written production provides an opportunity to replicate findings with orthographic lexical selection and segmental encoding processes which, presumably, should operate according to principles that are similar to those underlying phonological lexical selection and segmental encoding. Moreover, examining both modalities provides an opportunity to evaluate the robustness of findings across considerable task differences in terms of response execution: writing takes place over a longer time course and in a more serial fashion than speaking.

Interference and Facilitation in Language Production: Effects of Similarity

Having established which aspects of the language system are under investigation in this dissertation (lexical selection and segmental encoding in both the spoken and written modalities), one can consider the mechanisms underlying these stages of processing in more detail.

When a word is produced, both the target and its neighbors that share semantic and/or form-based features are activated. This results in a complex pattern of interference and facilitation arising from the dynamic nature of the processes involved in activation and selection. By *interference*, I mean reduced ease of production. A common usage of the term interference implies that there is direct competition: something interferes with something else. The term is used more broadly in this dissertation and in much of the past literature to describe effects that make production slower and/or less accurate, regardless of whether or not direct competition is involved. By *facilitation,* I mean increased ease of production. Again, this is a broad use of the term and does not imply anything about the cause of the facilitation effect. Facilitation can be observed in terms of speeded response times and/or increased accuracy. By manipulating the similarity of words that are to be produced, one can examine these interference and facilitation patterns and gain a better understanding of the underlying mechanisms of lexical selection and segmental encoding. In this chapter, I review two accounts relying on the same principles that have been proposed to explain the effects of semantic similarity on lexical selection. I then describe background work that I have previously conducted that extends these accounts beyond spoken production to the written modality (Breining & Rapp, submitted) and beyond lexical selection to segmental encoding (Breining, Nozari, & Rapp, 2016). These background studies provide a better understanding of how the mechanisms of the healthy production system function and will form a basis for hypotheses presented in later chapters that predict differences in learning as the result of different similarity contexts.

Interference and Facilitation in Lexical Selection

Although word production can be facilitated by presenting a single semantically related word immediately before the target, inserting an intervening item between them eliminates this benefit (e.g., Wheeldon & Monsell, 1994). Contrastingly, in paradigms in which items are named in the context of multiple similar items, robust and long-lasting interference and facilitation effects have been reported consistently. In the continuous naming paradigm that interleaves semantically related and unrelated pictures, participants are slower to name each consecutive item from the same semantic category. This occurs even when many unrelated items intervene between exemplars of the category (e.g., Brown, 1981; Damian & Als, 2005; Howard, Nickels, Coltheart, & Cole-Virtue, 2006; see Schnur, 2014 for data on the limitations of the effect). In the blocked cyclic naming paradigm, where small sets of pictures are named repeatedly, participants are slower to name the same items when they are in sets with others from the same semantic category (homogeneous blocks) than when they are in sets with unrelated items (heterogeneous blocks) (e.g., Belke, Meyer, & Damian, 2005; Breining et al., 2016; Crowther & Martin, 2014; Damian, Vigliocco, & Levelt, 2001; Kroll & Stewart, 1994; Schnur, Schwartz, Brecher, & Hodgson, 2006). These interference effects have been localized to lexical selection. The interference effect does not arise pre-lexically. When the same stimuli are presented as in blocked cyclic naming but the task is changed from naming to making non-verbal semantic judgments, responses are not slowed in cycles containing semantically similar pictures (e.g., Damian et al., 2001). Since the pre-lexical processing in this task is the same as in blocked cyclic naming but the interference effect disappears, the interference effect must arise after this point: it does not have a pre-lexical locus. Furthermore, the interference effect does not arise post-lexically. This is indicated by the finding that when the same stimuli are again presented but the task is changed to reading aloud, again the interference effect is not observed (e.g., Belke, Brysbaert, Meyer, & Ghyselinck, 2005; Belke, 2008, 2013; Damian et al., 2001). Reading aloud does not require lexical-semantic processing, but it does require the same post-lexical production of selected phonemes as naming aloud. Since the post-lexical processing is the same as in the naming task but the interference effect is not observed, the interference effect must arise prior to this point in processing: it does not have a post-lexical locus. Together, these findings suggest that the semantic interference effect arises at lexical selection, not pre-lexically or post-lexically.

Accounts of semantic interference. Two main accounts have been proposed that explain the long-lasting semantic interference effect observed in blocked cyclic naming (Howard et al., 2006; Oppenheim et al., 2010). Both accounts rely on the principles of shared activation, priming, and competition. *Shared activation* means that activation of a target results in the activation of related items as well. In lexical selection, semantic features send activation both to the target lexical node and to other lexical nodes that have those semantic features. *Priming* means that selection a target makes it more available on future trials as a target or competitor. After selecting a correct target, connections between that lexical node and the semantic features that activated it are strengthened. As a result, that target is more highly activated on future trials, both when it is again the target and when it is a competitor activated by related targets. *Competition* means that a target word is selected in the context of multiple active entries that are also available for selection. However, the specific implementation of these principles differs between proposals, as is illustrated in the following example. In the work presented here, I will not be adjudicating between the two models. The examples below show that both can account for the same interference effect by relying on the same principles of shared activation, priming, and competition, although there are some differences in how they are implemented.

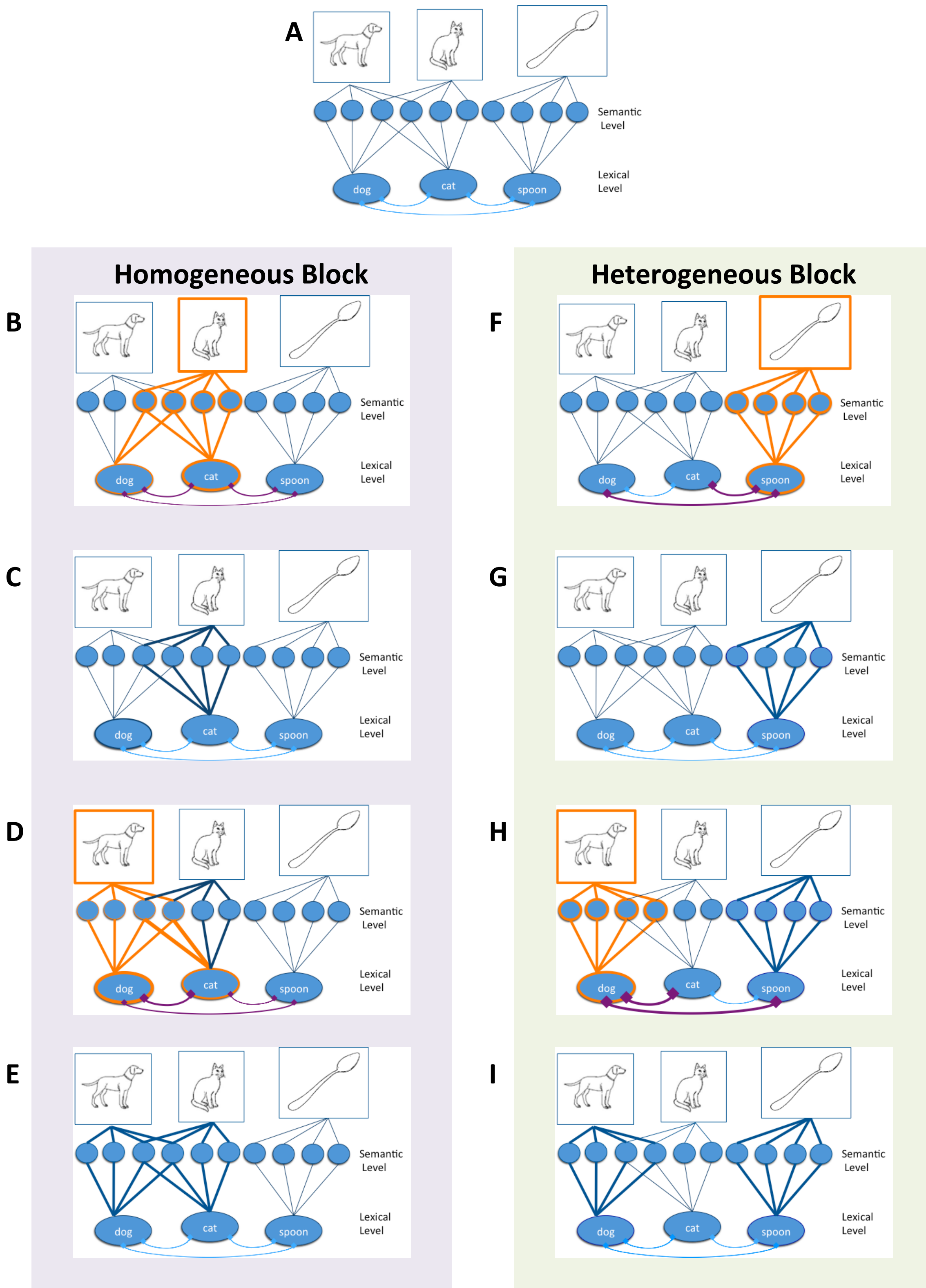
Howard and colleagues (2006) account. The first model, proposed by Howard and colleagues (2006), relies on competition via lateral inhibition between items at the lexical level. According to this model, when a picture is shown, its semantic features send activation to all lexical nodes that have those features (shared activation)[2]. The active lexical nodes inhibit one another in proportion to their own activation, which is a mechanism referred to as lateral inhibition (competition). In order for a lexical node to be selected, it must be more active than its competitors. Once a lexical node is selected, connections between that node and its semantic features are strengthened (i.e., the weights of the connections are increased), making that lexical node more available for selection on future trials when those semantic features are again activated (priming). Figure 3 depicts an example of this model.

Figure 3. Example implementation of the Howard and colleagues (2006) model of lexical selection.

In panel A, the initial state of the system is shown: there are three pictures, with connections to ten semantic features and then to three lexical nodes. Two of the lexical nodes, dog and cat, are semantically related, each sharing two of their four semantic features. Panels B-E, in the purple box, depict naming in a semantically homogeneous block. Here, the semantically related targets *cat* and then *dog* are named. Panels F-I, in the green box, depict naming in a semantically heterogeneous block. Here, the

[2] Note that there may be feedback from the lexical nodes to the semantic features. The existence of such feedback has been debated in the literature (for review and arguments against it, see Rapp and Goldrick, 2000). The presence or absence of such feedback does not substantively change this account or the later described Oppenheim and colleagues (2010) account. If such feedback is present in the system, it will simply increase the shared activation: active lexical nodes send feedback to their semantic features, which will in turn send additional activation to those lexical nodes and those of related items that share those features. For simplicity, this feedback is not included in the examples.

semantically unrelated targets *spoon* and then *dog* are named. Orange lines depict activation during each trial. Purple lines between lexical nodes depict lateral inhibition. Blue lines depict the connections between levels. The weight of these lines is used to depict strengthening, with thicker, solid lines showing strengthened connections. Refer to the text for a detailed description of the example depicted in this figure.


A
Semantic Level
Lexical Level
dog
cat
spoon
Homogeneous Block
Heterogeneous Block
B
C
D
E
F
G
H
I

In panel A, the initial state of the system is shown: there are three pictures, with connections to ten semantic features[3] and then to three lexical nodes. Two of the lexical nodes, *dog* and *cat*, are semantically related, each sharing two of their four semantic features. The additional lexical node, *spoon*, is not semantically related to the other lexical nodes, and does not share any of its semantic features with either of them.

Panels B-E illustrate what happens in a semantically homogeneous block in which the related pictures *cat* and *dog* are named.

In panel B, a picture of the first target, *cat*, is shown. This leads to activation of cat's semantic features, which in turn send activation to the lexical node *cat*. *Cat* and *dog* share many semantic features (e.g., *furry, pet*), so when a picture of a cat is shown, the lexical node *dog* is also partially activated. Lateral inhibition at the lexical level means that the active lexical nodes inhibit each other in proportion to their own activation. *Cat* is more active than *dog* because it receives activation from more semantic features: the distinctive features of *cat* that are not shared with *dog* (e.g., *meows*) contribute only to the activation of *cat*, not to the activation of *dog*. Therefore, *cat* inhibits *dog* more strongly than *dog* inhibits *cat*, and the lexical node *cat* can be selected since it is more active than all other competitors.

Panel C shows that after the lexical node *cat* is selected, the connections between that lexical node and its semantic features are strengthened. Other connections between semantic features and non-targets are unchanged.

---

[3]While the description I present here utilizes decomposed/distributed semantic representations, Howard and colleagues (2006) also simulated the same cumulative semantic interference effects in a model with non-decomposed/localist representations. The important point is that there is shared activation for semantically related items, regardless of whether that shared activation arises from activating nodes for semantic features that together make up the semantic representation of an item, from direct connections between nodes representing the whole meaning of an item, or from indirect connections between nodes that represent the whole meaning of a concept as part of a conceptual network.

Panel D shows what happens when the target of the next trial, *dog*, is semantically related to the target of the first trial, *cat*. Activating the semantic features of *dog* partially activates the previous response *cat* since they share semantic features. Because the connections between those shared features and the lexical node *cat* were strengthened on the previous trial, *cat* has greater activation and it inhibits *dog* more strongly than it would have had it not been the previous response. Therefore, there is interference: it takes longer to select *dog* after producing *cat* and selection may be more error-prone (as compared to a situation in which the previous target was not a related word) because *cat* is active and inhibiting *dog*.

Panel E shows that after selection of the second target *dog*, there is again strengthening of connections between the selected lexical node and the semantic features that contributed to its activation.

In contrast, panels F-I illustrate what happens in a semantically heterogeneous block in which the unrelated pictures *spoon* and *dog* are named.

In panel F, a picture of the first target, *spoon*, is shown. As in the previous example, its semantic features are activated and in turn send activation to the appropriate lexical node. However, in contrast to naming *cat* in the previous example, there are no items present that share semantic features with *spoon*. Therefore, only the lexical node for *spoon* receives activation. Since the features of *spoon* do not activate *dog*, there is no substantial lateral inhibition between the two.

Panel G shows that after selection of the first target *spoon*, there is strengthening of connections between the selected lexical node and the semantic features that contributed to its activation. Other connections between semantic features and non-targets are unchanged.

Panel H shows what happens when the target of the next trial, *dog*, is semantically unrelated to the target of the first trial, *spoon*. Here, activation of *dog* is unaffected by the previous trial. Since there is not shared activation since *dog* and *spoon* do not share semantic features, the strengthening of connections between the lexical node *spoon* and its features on the previous trial does not impact the activation of the lexical node *dog* from its semantic features. In contrast to the situation presented in panel D, there is not interference between these items: response times will not be slowed or accuracy reduced when *dog* is named after *spoon*.

Panel I shows that after selection of the second target *dog*, there is strengthening of connections between the selected lexical node and the semantic features that contributed to its activation. This strengthening does not interact with the previous strengthening of *spoon's* connections.

Thus, according to the Howard and colleagues (2006) model there is greater interference in a context where semantically related words are named (e.g., *dog* after *cat*; Panels B-E) than in a context where unrelated words are named (e.g., *dog* after *spoon*; Panels F-I). This interference is manifested in slower response times and/or reduced accuracy for naming responses in semantically homogeneous blocks as compared to semantically heterogeneous blocks.

Oppenheim and colleagues (2010) account. The second model, proposed by Oppenheim and colleagues (2010), does not rely on competitive selection via lateral inhibition but rather assumes that learning itself is a competitive process that both strengthens and weakens connections between representations[4]. Again, there is shared activation such that semantic features activate not only the target lexical node but also related competitors that have those features as well. Upon successful naming of a picture, connections between the correct lexical node and its semantic features are strengthened (i.e., the weights are increased meaning they carry greater activation), while connections between other lexical nodes that are active on that trial but not selected and the semantic features that activated them are weakened (i.e., the weights are decreased meaning they carry less activation) (competitive learning). Strengthened connections pass on greater activation, making it easier to select a lexical node when the semantic features connected to it by these strengthened connections are activated in the future, both when it is again the target or when it is a competitor (priming). Weakened connections pass on less activation, making it more difficult to selected a lexical node when the semantic features connected to it by these weakened connections are activated in the future, both when it is again a competitor or when it is the target. Figure 4 depicts an example of this model.

[4] These adjustments to connection weights are implemented in accordance with the delta rule tailored for the logistic activation function:

$$\Delta w_{ij} = \eta(a_i(1 - a_i)(d_i - a_i))a_j$$

Here, Δw_{ij} is the weight change for the connection to node *i* from node *j*; η is the learning rate; d_i is the desired activation of node *i*; a_i is the actual activation of node *i*; and a_j is the actual activation of node *j*. Activation ranges from 0 to 1. The learning rate η adjusts how quickly weight changes occur. Weight adjustments are scaled to a_i by the inclusion of $a_i(1 - a_i)$, meaning that the largest weight changes happen for moderate activations, not for those that are close to 0 or 1. Because $a_j$is included, the connections from node *j* to node *i* are only changed to the extent that node *j* is activated. Critically, connections are adjusted according to the discrepancy between the desired activation of the receiving node and its actual activation $(d_i - a_i)$: the weights are strengthened when there is less actual activation than desired, but they are weakened when there is more actual activation than desired.

Figure 4. Example implementation of the Oppenheim and colleagues (2010) model of lexical selection.

In panel A, the initial state of the system is shown: there are three pictures, with connections to ten semantic features and then to three lexical nodes. Two of the lexical nodes, *dog* and *cat*, are semantically related, each sharing two of their four semantic features. Panels B-E, in the purple box, depict naming in a semantically homogeneous block. Here, the semantically related targets *cat* and then *dog* are named. Panels F-I, in the green box, depict naming in a semantically heterogeneous block. Here, the semantically unrelated targets *spoon* and then *dog* are named. Orange lines depict activation during each trial. Blue lines depict the connections between levels. The weight of these lines is used to depict strengthening and weakening, with thicker, solid lines showing strengthened connections, and thinner, dashed lines showing weakened connections. Refer to the text for a detailed description of the example depicted in this figure.

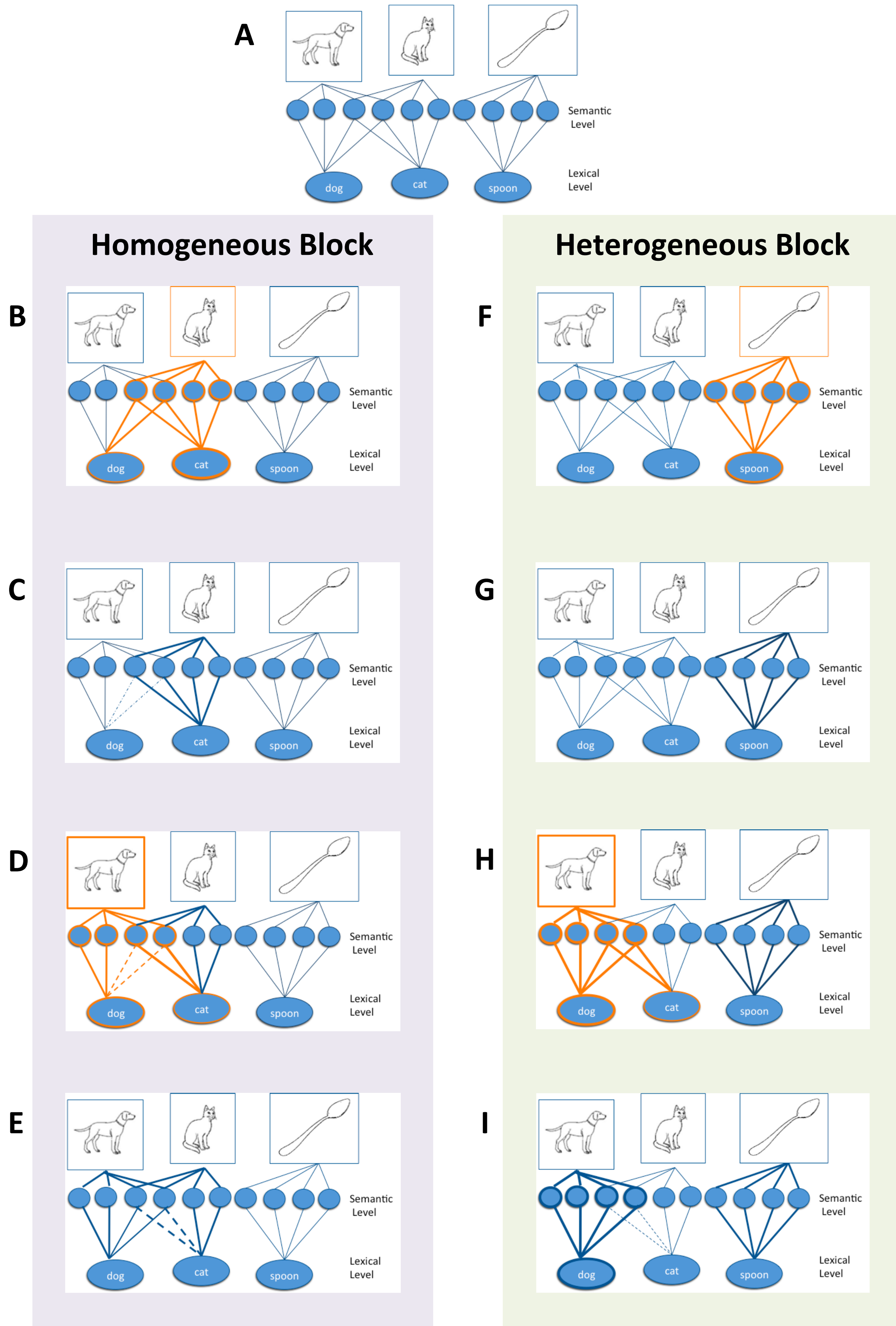

A
Semantic Level
Lexical Level
dog
cat
spoon
Homogeneous Block
Heterogeneous Block
B
C
D
E
F
G
H
I

Figure 4 has the same basic structure as Figure 3, again depicting a homogeneous block in which *cat* and *dog* are named and a heterogeneous block in which *spoon* and *dog* are named. In panel A, the initial state of the system is shown: there are three pictures, with connections to ten semantic features and then to three lexical nodes. Two of the lexical nodes, dog and cat, are semantically related, each sharing two of their four semantic features. The additional lexical node, *spoon*, is not semantically related to the other lexical nodes, and does not share any of its semantic features with either of them. Unlike Figure 3, there are not lateral inhibition connections between the lexical nodes.

Panels B-E illustrate a semantically homogeneous block in which *cat* and *dog* are named.

In panel B, a picture of the first target, *cat*, is shown. This leads to activation of *cat*'s semantic features, which in turn send activation to the lexical node *cat*. As in Figure 3, there will again be partial activation of related words that share semantic features, such as *dog*. As before, the distinctive features of *cat* contribute only to the activation of *cat*, not to the activation of *dog.* In this model there is no lateral inhibition. There is no direct competition between lexical nodes; rather, the lexical node that reaches the threshold of activation necessary for selection first is selected. In this case, the lexical node *cat* is selected: it receives activation from more semantic features and reaches the selection threshold first.

Panel C shows that after the lexical node *cat* is selected, the connections between its semantic features (e.g., *furry, pet, meows*) and its lexical node are strengthened. At the same time, the connections between the non-target related lexical node *dog* that was also activated by shared features and those shared features (e.g., *furry, pet*) are weakened.

The strengthening and weakening of connections is the implementation of competition in this model, as opposed to the lateral inhibition that implemented competition in the Howard and colleagues (2006) model described above.

Panel D shows what happens when the target of the next trial, *dog*, is semantically related to the target of the first trial, *cat*. Activating the semantic features of *dog* also leads to activation of *cat* through shared features. Because the connections between those shared features and the lexical node *cat* were strengthened on the previous trial, *cat* has greater activation than it would have had if it not been the previous response. *Dog* has less activation than it would have had if *cat* had not been the previous target because the connections between shared semantic features and dog were weakened on the previous trial. This reduced activation for *dog* means that there is interference: it takes longer for *dog* to reach the selection threshold in this situation as compared to one in which the previous target was not the related *cat*.

Panel E shows that after selection of the second target *dog*, there is again strengthening of connections between the selected lexical node and the semantic features that contributed to its activation, but weakening of connections between the non-target lexical node and the semantic features that contributed to its activation.

Contrast panels B-E with panels F-I, which illustrate what happens in a semantically heterogeneous block in which *spoon* and *dog* are named.

Panel F shows that when a picture of the first target, *spoon*, is shown, its semantic features are activated and in turn send activation to its lexical node. Because there are no items in this example that share semantic features with *spoon*, only that lexical node receives activation.

Panel G shows that after selection of the first target *spoon*, there is strengthening of connections between the selected lexical node and the semantic features that contributed to its activation. Since there was no activation of related items, no connections are weakened.

Panel H shows what happens when the target of the next trial, *dog*, is semantically unrelated to the target of the first trial, *spoon*. Here, activation of *dog* is unaffected by the previous trial. Since *dog* and *spoon* do not share semantic features, the strengthened connections between spoon's semantic features and lexical node are not involved and there was no previous modification of *dog*'s connections. In contrast to the situation presented in panel D, there is no interference here: naming of *dog* is not affected by the experience of previously naming *spoon*.

Panel I shows that after selection of the second target *dog*, there is strengthening of connections between the selected lexical node and the semantic features that contributed to its activation but weakening of connections between the non-target lexical node and the semantic features that contributed to its activation (e.g., *cat*'s connections are weakened even though *cat* is not named in this block). This strengthening and weakening does not interact with the previous strengthening of *spoon*'s connections.

Thus, according to the Oppenheim and colleagues (2010) model there is greater interference in a context where semantically related words are named (e.g., *dog* after *cat*) than in a context where unrelated words are named (e.g., *dog* after *spoon*). This interference has consequences for behavior, which can be observed as slower response times and reduced accuracy in the semantically homogeneous context as opposed to the semantically heterogeneous context. Although competition is implemented differently, this model predicts the same negative impact on the speed and accuracy of naming items in semantically related vs. unrelated blocks as the Howard and colleagues (2006) model.

On the two incremental learning accounts. Overall, both the Howard and colleagues (2006) and the Oppenheim and colleagues (2010) accounts agree that there is shared activation due to semantic similarity, that there are long-lasting changes to connection weights (i.e., learning) as a result of production, and that there is competition in the production system. As shorthand for these similarities, I will refer to both as incremental learning models. The primary difference between the accounts is how competition is implemented: the Howard and colleagues account relies on competitive lexical selection implemented as lateral inhibition, while the Oppenheim and colleagues account relies on a competitive learning process implemented as increasing and decreasing of connection weights via the delta rule. The background studies I present in this chapter will not allow us to distinguish between these accounts since both models predict the same behavioral interference effect will result from similarity. However, some have argued that the Oppenheim incremental learning model may be preferred because it does not rely on competitive selection to account for cumulative semantic interference (see Navarrete, Del Prato, Peressotti, & Mahon, 2014 for arguments against competitive selection). For simplicity, in the rest of the dissertation I will consider extensions of the Oppenheim and colleagues (2010) model. Extensions of the Howard and colleagues (2006) model lead to compatible predictions, but will not be discussed in detail. The work presented here will test the predictions of incremental learning models in general, but will not adjudicate between different implementations of these models. It is important to note that at present there is not a clear alternative to the incremental learning models proposed to explain the same effects (chiefly long-lasting semantic interference). The work presented here will be confirmatory in the sense that it will test whether or not the predictions of incremental learning models and their extensions hold; it will not be a direct comparison pitting alternative accounts against one another.

Background experiment 1: Semantic similarity in written word production. The Howard and colleagues (2006) and the Oppenheim and colleagues (2010) incremental learning models were proposed to account for semantic interference effects in spoken production. Can they also account for effects when the modality is instead written production? Prior to the background work presented here, this question had not been addressed. As discussed previously, written and spoken production are independent modalities that abide by a set of general cognitive principles. If the mechanisms proposed in the two accounts do in fact instantiate domain-general cognitive principles, the semantic interference effect should be replicated when the modality of production is written.

Method. In order to address these questions, Background Experiment 1 was performed (alse reported in Breining & Rapp, submitted). This experiment sought to replicate the semantic blocking effect in a written blocked cyclic picture naming paradigm, extending previous work to written production. The blocked cyclic naming paradigm was chosen for a number of reasons. First, it parallels the production studies that have consistently reported interference as a result of semantic overlap. This paradigm was preferred to continuous naming because of its applicability to clinical and educational settings, both of which often involve repeated presentation of items that may be grouped based on either meaning or form. If practicing items with high similarity creates interference during training, this finding may have important consequences for long-term learning. Such consequences are the focus of this dissertation. A picture-naming task was chosen because it requires use of a lexical route that necessarily includes both lexical selection and segmental encoding. This means the same task could be used to investigate questions about both processes (see other background experiments reported in this chapter). It additionally avoids the confound present in written spelling to dictation whereby some participants may use predominantly sublexical or non-semantically-mediated processes to perform the task.

The paradigm used in Background Experiment 1 was identical to those that have been reported previously (e.g., Belke, Meyer, & Damian, 2005; Crowther & Martin, 2014; Damian, Vigliocco, & Levelt, 2001; Kroll & Stewart, 1994; Schnur, Schwartz, Brecher, & Hodgson, 2006), except that naming was performed via writing instead of speaking. Twenty undergraduate participants named pictures as quickly and accurately as possible by writing responses on a graphics tablet. In total, there were thirty-six black and white pictures selected from those used by Schnur and colleagues (2006), consisting of six exemplars of six categories (animals, body parts, clothes, furniture, toys, eating utensils). The same pictures appeared in the context of related items (homogeneneous blocks) and in the context of unrelated items (heterogeneous blocks). Heterogeneous and homogeneous blocks were alternated in pseudorandom order, with periodic breaks. Within each block, there were four presentations (cycles) of the six pictures in random order. According to the models of lexical selection described above, assuming that written production shares the same principles as spoken production, reliable interference should be observed for items in homogeneous blocks as compared to heterogeneous blocks.

Table 1. Items used in Background Experiment 1. The six items in each row form a homogeneous block, with items drawn from the same semantic category. The six items in each column form a heterogeneous block, with items drawn from different semantic categories.

	heterogeneous blocks					
homogeneous blocks	bear	cat	skunk	goat	horse	dog
	chin	nose	ear	toe	thumb	arm
	hat	glove	dress	sock	coat	skirt
	sofa	bed	table	crib	stool	chair
	doll	top	bat	ball	blocks	kite
	glass	spoon	cup	knife	pitcher	fork

Analysis. Data analysis focused on response time measures after removal of error responses and outliers greater than 2.5 standard deviations from each participant's overall mean. Only the second through fourth cycles were considered as past reports indicate that block type effects typically emerge only after the first cycle: they initially may be obscured by facilitation due to repetition priming (Belke, Meyer, et al., 2005). Repeated measures 2x3 Analysis of Variance (ANOVA) included block type (homogeneous or heterogeneous) and cycle (2,3,4) as within-subject factors. This same analysis was used for all four background experiments reported in this chapter.

Results and discussion. The results of Background Experiment 1 were consistent with previous reports of semantic interference in spoken production. Critically, participants were slower to initiate production of items in homogeneous blocks than heterogeneous blocks. There was no significant main effect of cycle or interaction of cycle and block type[5] (see Figure 5).

This experiment replicated the semantic interference effect previously reported in spoken production in the domain of writing, providing evidence consistent with the hypothesis of similar principles underlying lexical selection in both modalities. These mechanisms are compatible with the models proposed by Howard et al. (2006) and Oppenheim et al. (2010). Further support for this conclusion can be drawn from a similar study conducted by Nozari and colleagues (2016), which found semantic interference for both written and spoken picture naming using a slightly different paradigm in which semantically-related items were repeatedly named in two-item blocks.

---

[5] The Oppenheim and colleagues (2010) model predicts cumulative semantic interference that continues to grow across cycles, which should be visible through the interaction of cycle and block type. However, a significant interaction has not consistently been found in empirical studies of blocked cyclic naming (for review, see Belke & Stielow, 2013). The contrast in observed vs. predicted effects may reflect differences in the exact balance of interference due to blocking and facilitation due to repetition priming when repeatedly producing the same word, may demonstrate contributions of language-external control mechanisms, or may be due to insufficient power to detect such effects in the individual empirical studies. Further investigation is necessary. However, it is not the interaction between cycle and block type that is of chief concern in testing whether blocking leads to interference; rather, the critical effect is that of block type. Observing interference in homogeneous as opposed to heterogeneous blocks supports the idea that similarity leads to long-lasting interference, regardless of whether or not this interference continues to grow over cycles.

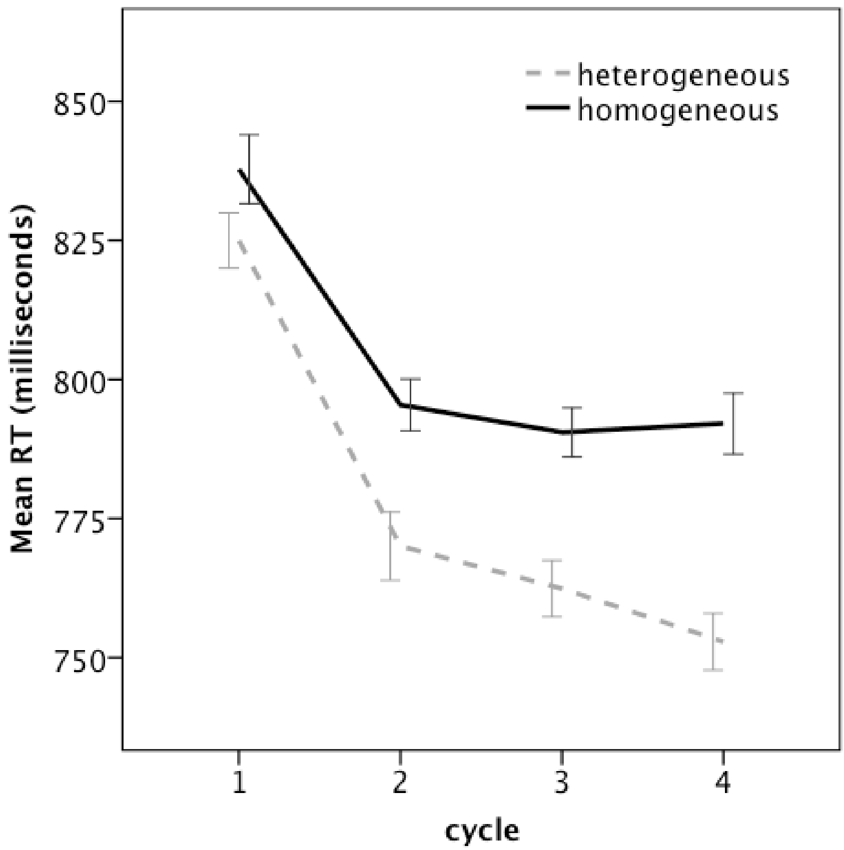


Figure 5. Results of Background Experiment 1: The effect of semantic similarity on response times for written picture naming of words in semantically homogeneous versus semantically heterogeneous blocks.
Error bars represent the between-subjects standard errors of the means, corrected for repeated measures using the Cousineau (2005) method.

Interference and Facilitation in Segmental Encoding

Having observed evidence of interference due to semantic similarity consistent with the models of lexical selection described above, I now turn to segmental encoding, another key process of language production. Although incremental learning accounts have primarily focused on semantic-lexical mapping, they may also be applicable to lexical-segmental mapping. Lexical-segmental mapping, like semantic-lexical mapping, is affected by learning. As one example of incremental learning applied to this stage of processing, Mulatti and colleagues (2012) reported slowed response latencies when participants read aloud words that rhymed with previously produced words. However, this observed interference could potentially have arisen in sublexical grapheme-to-phoneme mapping instead of through the lexical processes that are of interest here since the study was conducted in Italian, an orthographically transparent language in which graphemes can be reliably mapped to phonemes.

Extending an incremental learning model to segmental encoding. How might incremental learning work in lexical-segmental mapping? The example below extends the Oppenheim and colleagues (2010) model from lexical selection to segmental encoding, in line with the extension suggested by Breining, Nozari, and Rapp (2016). For completeness, Appendix A describes a similar extension of the Howard and colleagues (2006) account. Note that both accounts lead to the same predictions regarding the behavioral effects of producing words in segmentally related vs. unrelated contexts.

The logic of Oppenheim and colleague's (2010) model can be applied to address the segmental encoding stage of production. The same principles of shared activation, priming, and competitive learning apply as in the account of lexical selection. After a lexical node is activated, activation is sent to the constituent segments that make up that word. Shared activation results from feedback between the segmental and lexical levels (e.g., Dell et al., 2014; Rapp & Goldrick, 2000). Activated segments send feedback not only to the lexical node that activated them, but also other lexical nodes that share the activated segments. Thus lexical nodes that share segments are activated as a target is processed. Priming and competitive learning are again implemented as strengthening and weakening of connections. When a word is produced, the connections between the target segments and the lexical nodes that contribute to their activation are strengthened in proportion to their activation level, while the connections between non-target segments and the lexical nodes that contribute to their activation are weakened in proportion to their activation level. Critically, due to shared activation via feedback, these other active non-target lexical nodes are likely to be those that share segments with the target. Figure 6 depicts an example of this model.

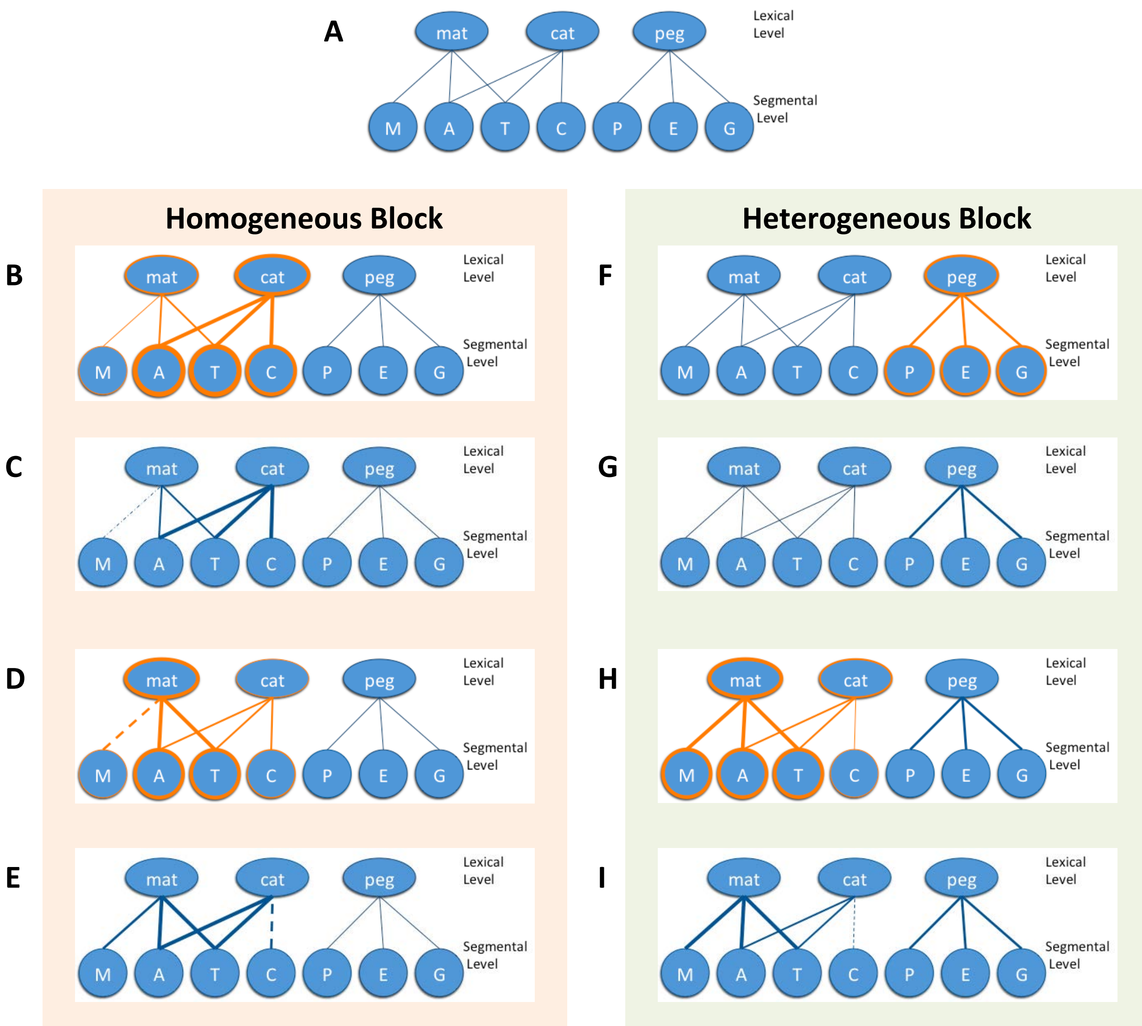


Figure 6. Example implementation of an extended version of the Oppenheim and colleagues (2010) model that applies to segmental encoding.

In panel A, the initial state of the system is shown: there are three lexical nodes, with connections to seven segments. Two of the lexical nodes, *mat* and *cat*, are segmentally related, each sharing two of their three segments, while the third, *peg*, is not. Panels B-E, in the orange box, depict naming in a segmentally homogeneous block. Here, the segmentally related targets *cat* and then *mat* are named. Panels F-I, in the green box, depict naming in a segmentally heterogeneous block. Here, the segmentally unrelated targets *peg* and then *mat* are named. Orange lines depict activation during each trial. Blue lines depict the connections between levels. The weight of these lines is used to depict strengthening and weakening, with thicker, solid lines showing strengthened connections, and thinner, dashed lines showing weakened connections. Refer to the text for a detailed description of the example depicted in this figure.

Here, instead of the semantic and lexical levels, the lexical and segmental levels are depicted. Panel A shows the initial state of the system: there are three lexical nodes, with connections to seven segments[6]. Two of the lexical nodes, *mat* and *cat*, are segmentally related, each sharing two of their three segments. The third lexical node, *peg,* is not connected to any of the same segments as the other two.

Panels B-E illustrate what happens in a segmentally homogeneous block in which *cat* and *mat* are named.

In panel B, the lexical node *cat* is activated. This leads to activation of *cat*'s segments C, A, and T. Through feedback, these segments send activation to the lexical node *mat* that shares many of the same segments. *Mat* in turn sends activation to its segments, including M that is not part of the target *cat*. Overall, *cat* is more active than *mat*, and its segments are selected.

Panel C shows that when the correct target *cat* is produced, the connections that contributed to activation of its segments are strengthened. These strengthened connections include the ones between the target segments A and T and the non-target lexical node *mat* that shares them because all connections that contribute to activation of the correct target segments are strengthened. At the same time, active connections between unshared, non-target segments and the lexical nodes that contributed to their activation are weakened. This includes the connection between M and *mat*, which was activated via feedback.

[6] In this simplified model, only information about the identity of segments is shown at the segment level. Information about the position of these segments is also likely to be represented at the segmental level. Additionally, the segments shown here happen to be graphemes, although the same logic applies when segments are instead phonemes.

Panel D shows what happens when the target of the next trial, *mat*, is segmentally related to the target of the first trial, *cat*. Activating the segments of *mat* also leads to the activation of *cat* through feedback. Because the connections between the lexical node *cat* and the shared segments A and T were strengthened on the previous trial, *cat* has greater activation than it would have had had it not been the previous response. *Mat*, meanwhile, has less activation than it would have had had *cat* (as opposed to an unrelated word) not been the previous target because the connection between *mat* and M was weakened on the previous trial. This reduced activation for *mat* means that there is interference: it takes longer for *mat* to reach the selection threshold in this situation as compared to one in which the previous target was not the related *cat*[7].

Panel E shows that after production of the second target *mat*, there is again strengthening of connections between the selected segments and the lexical nodes that contributed to their activation, but weakening of connections between the non-target segments and non-target lexical nodes that were activated via feedback.

In contrast, panels F-I illustrate a segmentally heterogeneous block in which *peg* and *mat* are named.

Panel F shows that when the lexical node of the first target, *peg*, is activated, it sends activation to its segments. Because there are no items in this example that share segments with *peg*, there is no feedback that leads to activation of non-target lexical nodes or segments.

---

[7] This discussion makes the assumption that the speed of production will be determined by the slowest selection of a segment; the weakest connection is determinative in this situation. Thus the weakening of the connection between the unshared segment and lexical node on the previous trial leads to interference when that item becomes the target, even though there was also strengthening of connections between the shared segments and lexical node of the new target on the previous trial in which a segmentally related word was the target.

Panel G shows that when the correct target *peg* is produced, the connections that contributed to activation of its segments are strengthened. Since there were no active connections that contributed to activation of non-target segments, no connections are weakened in this example (if non-target segments had been activated, their connections to lexical nodes would be weakened).

Panel H shows what happens when the target of the next trial, *mat*, is segmentally unrelated to the target of the first trial, *peg*. Here, activation of *mat* is unaffected by the previous trial. Since *mat* and *peg* do not share segments, the strengthened connections between peg's lexical node and segments are not involved, and there was no previous modification of mat's connections. In contrast to the situation presented in panel D, there is no interference in this situation.

Panel I shows that after production of the second target *mat*, there is strengthening of connections between its segments and the lexical nodes that contributed to their activation (M A T and *mat*; A T and *cat*), but weakening of connections between non-target segments and the lexical nodes that contributed to their activation (C and *cat*). This strengthening and weakening does not interact with the previous strengthening of *peg*'s connections since neither the lexical node *peg* nor any of its segments are activated.

According to this extended model, interference is predicted when segmentally related words as opposed to segmentally unrelated words are produced. Segments of an unrelated word are much less likely to be activated and undergo weakening of connections to a target than are segments of related words. Therefore, when the second target is segmentally related to the first target, it is at a disadvantage compared to when it follows an unrelated first target. If this extended model is correct, the predicted interference should be visible in slowed response times and reduced accuracy for naming in segmentally related vs. unrelated contexts.

Differences in semantic-lexical and lexical-segmental mapping. Semantic-lexical and lexical-segmental mappings differ in important ways. Semantic-lexical mapping requires connecting *many* semantic features to *one* lexical node, while lexical-segmental mapping requires connecting *one* lexical node to *many* segments. According to the incremental learning models described above, both types of mappings involve strengthening connections that support correct activation of the target but not those that support incorrect activation of non-targets. In the semantic-lexical case, there is only one correct target lexical node, so only the connections from semantic features that contribute to it are strengthened. Because of shared activation, those same features also contribute to activation of non-targets words, but these non-target connections are not strengthened: according to the Howard and colleagues account they are unchanged, while according to the Oppenheim and colleagues account they are weakened. That is, connections from distinctive (unshared) semantic features to lexical nodes are strengthened, while connections from shared semantic features to lexical nodes are weakened.

On the other hand, in the lexical-segmental case, there are multiple correct target segments, and all the connections from lexical nodes that contribute to their activation are strengthened. Because of shared activation via feedback, it is not only the target lexical node that contributes to target segment activation, but also non-target lexical nodes that share those segments. According to incremental learning models, these connections between non-target lexical nodes and target segments are strengthened. According to the Oppenheim and colleagues account, the connections between non-target lexical nodes and non-target segments are weakened. That is, connections from lexical nodes to shared segments are strengthened, while connections from lexical nodes to distinctive (unshared) segments are weakened.

Therefore, while similarity leads to behaviorally observable interference in both types of mapping, the source of this interference that arises in later trials after related items have been named is different. There is interference in semantic-lexical mapping because shared semantic features contribute to activation of non-target lexical nodes (i.e., the activation of *shared* features makes selection of the correct lexical node slower and/or more error-prone). There is interference in the case of lexical-segmental mapping because the non-target lexical nodes of items that share many segments with the target contribute to activation of unshared segments (i.e., the activation of *distinctive/unshared* segments makes selection of the correct segments slower and/or more error-prone).

Consequences for shared and distinctive features. Over time these different sources of interference have different effects on the relative weightings of connections at the two levels as learning takes place. Learning takes the form of adjustments that decrease the strength of competitors relative to the current target word, and in turn reduce the availability of those competitors on subsequent trials where they become targets. This leads to observable interference. According to an incremental learning model analogous to that proposed by Oppenheim and colleagues (2010), when lexical selection proceeds in the context of semantically related items, the connections between shared semantic features and lexical nodes will be weakened, while the connections between distinctive (unshared) semantic features and lexical nodes will be strengthened (see Figure 4 panel E). In contrast, when segmental encoding proceeds in the context of segmentally related items, the connections between lexical nodes and shared segments will be strengthened, while the connections between lexical nodes and distinctive (unshared) segments will be weakened (see Figure 6 panel E)[8]. I will return to the differences between the consequences of overlap on lexical selection and segmental encoding in later chapters as they lead to different predictions about how different types of similarity in training sets may impact long-term learning.

Do incremental learning models apply to segmental encoding?

Past research. It is not clear that past empirical findings support the proposed incremental learning model for segmental encoding: both facilitation and interference effects have been reported as a result of producing words in contexts with high segmental similarity. Single-trial primes have yielded mixed results, with some studies reporting facilitation (e.g., Damian & Dumay, 2009) and some interference (e.g., Sullivan & Riffel, 1999). Likewise, studies of the effects of neighborhood density on word production have led to conflicting results, although careful analysis suggests predominately interference effects (for review, see Sadat, Martin, Costa, & Alario, 2014). Studies involving repetition of words pairs report interference for overlapping onsets and facilitation for overlapping rhymes (e.g., Sevald & Dell, 1994). However, repetition places minimal demands on lexical selection and can potentially occur sublexically. Facilitation effects

[8] A model analogous to that proposed by Howard and colleagues (2010) makes a similar prediction about segmental encoding in the presence of segmental overlap: the connections between lexical nodes and shared segments will be strengthened relative to the connections between lexical nodes and distinctive segments because the connections to shared segments are strengthened on every trial while the connections to distinctive segments are only strengthened when those segments are correct targets (see Figure A1 panel E). A model like that of Howard and colleagues makes less clear predictions about lexical selection. An implementation of the model that explicitly represents semantic features has no obvious reason to differently weight connections between shared features and lexical nodes vs. connections between distinctive features and lexical nodes since this model has only strengthening of connections to correct targets, not weakening of connections that contribute to non-target activation. However, a different implementation with non-decomposed representations as nodes in a conceptual network might make different predictions that align or diverge from those of the Oppenheim and colleagues model.

are commonly reported for paradigms such as associative cueing, in which participants learn pairs of words and then produce the second member of a pair in response to the first (e.g., Damian & Bowers, 2003; Meyer, 1990, 1991; Roelofs & Meyer, 1998; Roelofs, 1999), and the picture-word interference paradigm, in which participants name a picture while reading or hearing a different word (Costa & Caramazza, 2002; Meyer & Schriefers, 1991; Schriefers, Meyer, & Levelt, 1990).

Robust interference for form-related items has not been consistently reported in situations that typically produce interference for semantically related items, including the blocked cyclic naming paradigm. Indeed, when all or the majority of items in a cycle share the same onset segment, production is facilitated (e.g., Damian, 2003; Hodgson, Schwartz, Schnur, & Brecher, 2005; Roelofs, 1999; Schnur et al., 2009). However, shared onset facilitation has been widely attributed to the strategic preparation made possible by high predictability (e.g., Damian, 2003; Meyer, 1991). For example, if a participant realizes that all the items in a block begin with /b/, he or she can prepare to produce that sound before even seeing the picture on a given trial. This means that the participant may be able to start producing the initial sound before retrieving the rest of the segments, speeding the onset of responses, which is the typical measure analyzed in these studies. Strategic preparation, which likely arises outside the language production system (O'Séaghdha & Frazer, 2014), may mask interference effects generated within the system. There is some evidence generally consistent with the prediction that removing predictability reveals underlying interference due to segmental overlap. Belke and Meyer (2007) reported that facilitation effects on response latencies disappeared when multiple onset-related items were named within a single trial and that gaze durations to onset-related items increased when items were named quickly. However, overt interference effects comparable to those in semantically related naming paradigms had not been reported prior to the background studies reported here. This failure to find interference due to segmental overlap could mean that competitive incremental learning is not operational in segmental encoding, or that the predicted interference is masked by the strategic preparation that is possible in conditions with predictable initial segment overlap.

Background experiments. In the three background studies reported below, I investigated these possibilities. First, I replicated previous spoken blocked cyclic naming studies that reported facilitation when initial segments are shared in the written modality (Background Experiment 2). I then examined the consequences of segmental overlap when it is distributed unpredictably across positions in words as opposed to being limited to the first position (e.g., *pill* in the context of *pig, peg, pot, log, leg*), utilizing the same blocked cyclic naming paradigm in both spoken and written production (Background Experiments 3 and 4; recently reported in Breining, Nozari, & Rapp, 2016). This paradigm was applied for the same reasons as those described above for Background Experiment 1: it is parallel to the paradigm used to investigate semantic similarity, meaning lexical selection and segmental encoding could be examined using the same basic task; it is applicable to training in clinical and educational settings, which are a later part of this dissertation; and it requires use of the lexical route of processing as opposed to permitting reliance on the sublexical route. This last point is especially important for investigating the impact of segmental similarity on segmental encoding. While there is some evidence of segmental overlap interference in reading (e.g., Mulatti et al., 2012) and repetition (e.g., O'Séaghdha & Marin, 2000; Sevald & Dell, 1994), these results may be due to similarity on input or non-lexical processing, making such tasks less appropriate for the investigation of lexical selection and segmental encoding. Opportunities for strategic preparation were reduced in Background Experiments 3 and 4 by removing predictability, allowing evaluation of whether principles such as incremental learning operate during both lexical selection and segmental encoding.

These experiments investigated both spoken and written word production. As described above, extending the work to written production provided a replication opportunity. Using both modalities permitted direct investigation of the assumption that orthographic segmental encoding processes operate according to similar principles as phonological segmental encoding, while allowing examination of the robustness of findings across task differences such as the longer, more serial time course of written response execution.

Background experiment 2: Initial segment overlap in written word production. In this experiment, a written version of the blocked cyclic naming paradigm was employed to determine if effects of facilitation based on initial segment overlap are observed in written word production, congruent with those previously observed in spoken word production.

Method. The method was very similar to Background Experiment 1. Twenty-four undergraduates participated, none of whom had been part of Experiment 1. This time, the thirty-six stimuli were pictures corresponding to monosyllabic, consonant-vowel-consonant, 3-4-letter words. Within homogeneous lists, all words shared the first letter. The procedure and analysis were the same as in Background Experiment 1.

Table 2. Items used in Background Experiment 2. The six items in each row form a homogeneous block, with all items sharing the first letter. The six items in each column form a heterogeneous block, with no items sharing the first letter.

homogeneous blocks	heterogeneous blocks					
	bed	boat	book	bib	bug	bat
	coal	cup	cab	cog	coat	coin
	fan	fin	fig	foam	foil	foot
	mop	maid	mat	moon	man	mug
	root	rug	rain	rat	rib	road
	tub	tag	toad	team	tool	top

Results and discussion. The results of Background Experiment 2 were consistent with previous reports of facilitation for initial segment overlap in spoken word production. Critically, as depicted in Figure 7, participants were faster to initiate production of items in homogeneous blocks than in heterogeneous blocks. There was a significant effect of cycle, but the interaction between cycle and block type did not reach significance.

This experiment showed that the onset facilitation effect was robust to changes in modality, uncovering an effect in written production that is consistent with that reported for spoken production. This finding provided greater confidence that the blocked cyclic naming paradigm leads to reliable effects that are robust to changes in modality, which was important for the next two experiments that took advantage of a novel manipulation in the same paradigm. The significant effect of cycle is consistent with the idea that repetition priming played a role throughout this experiment that was not counteracted by interference.

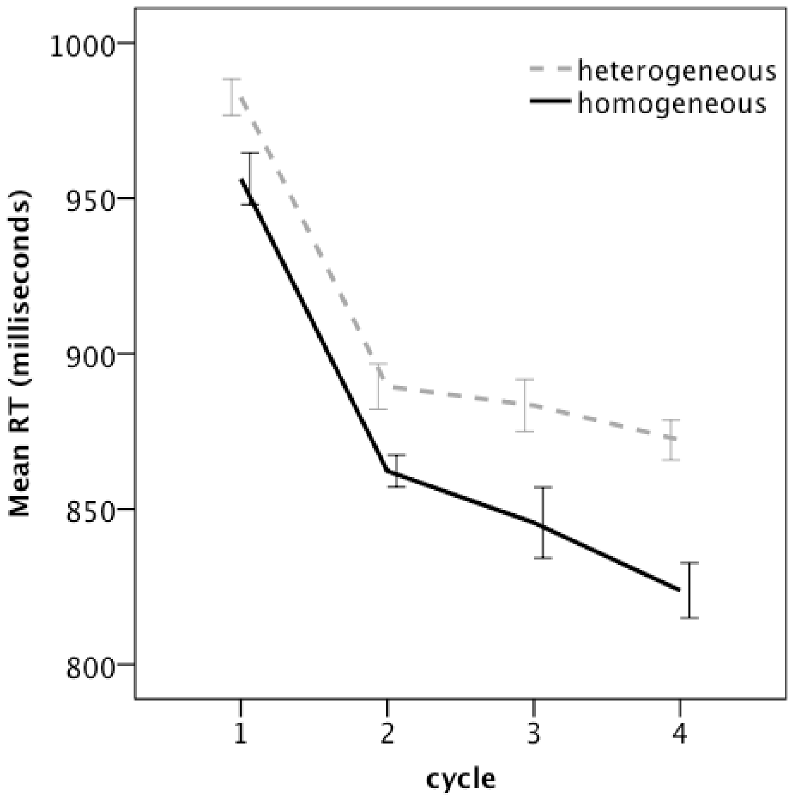


Figure 7. Results of Background Experiment 2: The effect initial segment overlap on response times for written picture naming of words in blocks in which words shared their initial segment (homogeneous) versus blocks in which they did not (heterogeneous). Error bars represent the between-subjects standard errors of the means, corrected for repeated measures using the Cousineau (2005) method.

Background experiment 3: Distributed segmental overlap in spoken word production. In Background Experiment 3, the effects of unpredictable segmental overlap were examined in a spoken blocked cyclic naming paradigm, investigating whether interference occurs in this situation. Such a finding would suggest that similar principles underlie lexical selection and segmental encoding.

Method. The spoken paradigm used was similar to previous reports utilizing blocked cyclic naming, but the similarity manipulation was novel. Twenty-four undergraduate participants were recruited. In this experiment, the stimuli consisted of thirty-six pictures corresponding to monosyllabic 3-6 letter words. In the homogeneous blocks words shared many phonemes, whereas in the heterogeneous blocks words shared few phonemes. This was quantified using a measure of position-independent phonological overlap, defined as the total number of phonemes shared by two strings, regardless of position, divided by the total number of phonemes in the two strings (Goldrick, Folk, & Rapp, 2010). Stimuli are listed in Table 1. Procedures were the same as in Background Experiments 1 and 2, except that participants did not write their responses, but rather responded by speaking into a microphone. The primary analysis was the same as in Background Experiments 1 and 2.

Table 3. Items used in Background Experiments 3 and 4. The six items in each row form a homogeneous block, with high position-independent segmental overlap. The six items in each column form a heterogeneous block, with low position-independent segmental overlap.

homogeneous blocks	heterogeneous blocks					
	cat	mat	cot	cap	map	mop
	pill	peg	pig	pot	log	leg
	house	horse	rose	nose	robe	hose
	rain	stairs	hair	stain	chain	chair
	slide	bride	bread	bridge	sled	bird
	belt	well	wall	bell	bull	ball

Results and discussion. The results indicated that, critically, participants were slower to initiate production of items in homogeneous blocks than in heterogeneous blocks. As in Background Experiment 1, there was no significant main effect of cycle or interaction of cycle and block type (Figure 8A). Thus, distributed segmental overlap resulted in interference in spoken production.

A secondary analysis included only the 21 items that shared their first segment with at least half the items in their homogeneous block in order to see if interference was observed even when items shared onsets, the condition past research suggests is most likely to yield facilitation in predictable situations. Again, participants were slower to initiate production in homogeneous blocks (Figure 8B). This indicates that interference is found when onsets overlap if predictability is eliminated, suggesting that past reports of facilitation were driven by strategic preparation made possible by high predictability.

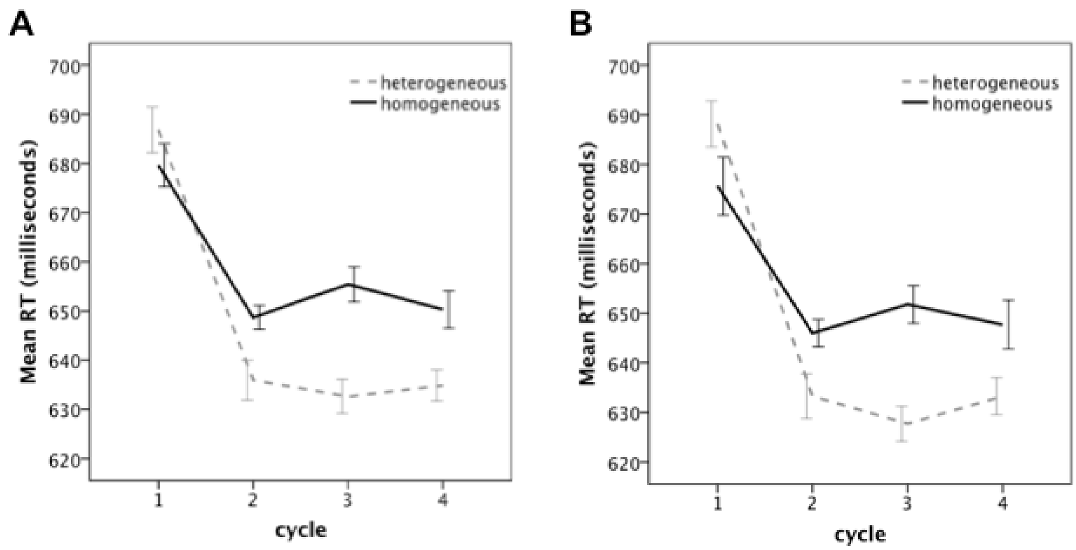


Figure 8. Results of Background Experiment 3: The effect of distributed segmental overlap on response times for spoken picture naming of words in blocks with high segmental overlap (homogeneous) versus low segmental overlap (heterogeneous). Error bars represent the between-subjects standard errors of the means, corrected for repeated measures using the Cousineau (2005) method. Panel A includes data from all items. Panel B depicts the secondary analysis that includes only items that share the initial segment with at least half of the other items in the homogeneous block.

Background experiment 4: Distributed segmental overlap in written word production. Background Experiment 4 tested whether the results of Background Experiment 3 were replicated in written production. Finding a similar interference effect would indicate that reliable segmental overlap interference occurs for both phoneme and letter segments.

Method. Thirty-four undergraduates participated in this experiment. The stimuli and analysis were the same as in Background Experiment 3; the procedures were the same as in Background Experiments 1 and 2. Evaluating heterogeneous vs. homogeneous blocks in terms of orthographic rather than phonological overlap verified that there was significantly greater position-independent letter overlap in the homogeneous blocks than the heterogeneous blocks.

Results and discussion. The primary analysis of all stimuli revealed a significant effect of block type (Figure 5A). Critically, participants were again slower to initiate production of items in homogeneous blocks than in heterogeneous blocks. Again, there was no significant main effect of cycle or interaction of cycle and block type. As in Background Experiment 3, these results point to interference induced by distributed segmental overlap.

In order to directly compare the results of Background Experiments 3 and 4, data from both experiments were entered into the same model, with modality as a between-subjects factor. The critical main effect of block type remained significant. However, the interaction between modality (written vs. spoken) and block type (heterogeneous vs. homogeneous) was not significant. This indicated that there were no reliable differences in interference found in spoken and written modalities.

A secondary analysis comparable to that of Background Experiment 3 was also conducted. As in that experiment, there was a numerical interference effect when only items sharing initial segments with at least half of the items in their homogeneous block were analyzed (Figure 5B). Although this effect did not reach statistical significance, the effect was not reliably different from that of Background Experiment 3, as indicated by the significant effect of block type but non-significant interaction of modality and block type in analysis of the combined data. Overall, there was comparable robust interference in both the spoken and written modalities for segmentally overlapping words.

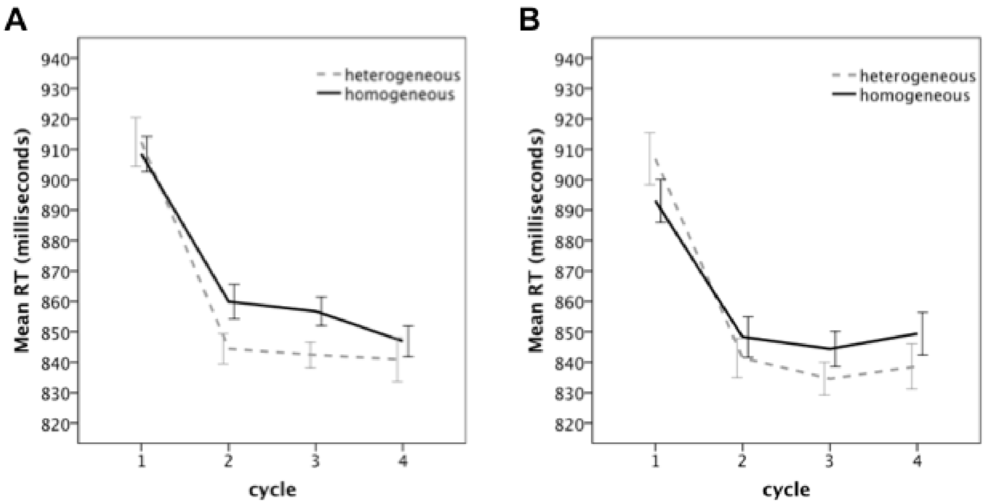


Figure 9. Results of Background Experiment 4: The effect of distributed segmental overlap on response times for written picture naming of words in blocks with high segmental overlap (homogeneous) versus low segmental overlap (heterogeneous). Error bars represent the between-subjects standard errors of the means, corrected for repeated measures using the Cousineau (2005) method. Panel A includes data from all items. Panel B depicts the secondary analysis that includes only items that share the initial segment with at least half of the other items in the homogeneous block.

General Discussion of Background Work

This chapter presented four background experiments that examined whether the consequences of overlap are similar across stages and modalities of word production using a blocked cyclic naming paradigm. In the first two experiments, the well-established consequences of representational overlap in spoken production were replicated in written production. In Background Experiment 1, participants were slower to initiate written responses for pictures in semantically homogenous blocks than in semantically heterogeneous blocks. In Background Experiment 2, participants were faster to initiate written responses when picture names shared the first segment than when they did not share segments. These experiments increase confidence that the facilitation and interference effects reflect similarity in the underlying mechanisms across modalities that are robust to changes in task conditions. The last two experiments used a novel similarity manipulation in which segmental overlap was distributed unpredictably across word positions. This manipulation led to interference in both spoken (Background Experiment 3) and written (Background Experiment 4) production, even when considering only items that shared their initial segment with at least half the items in their homogeneous block. Critically, the interference observed for items with distributed segmental overlap is parallel to the interference observed for items with semantic overlap (Background Experiment 1), not the facilitation found when picture names predictably shared onset segments (Background Experiment 2). The interference effect was replicated across modalities and was not reliably different between the two, which increases confidence that the effects are stable even across considerable variability in task conditions. Further support for these results can be drawn from a similar study conducted by Nozari and colleagues (2016) using repeated picture naming of two-item blocks in both the spoken and written modalities. This study reported comparable effects of facilitation when the words in a block predictably shared initial segments and of interference when they predictably shared final segments.

Implications for Theories of Word Production. The results of the background experiments have several implications for word production theories.

Interference at both stages of production. First, there is evidence that similarity-based interference occurs for both the lexical selection and segmental encoding stages. In general, distributed feature overlap creates interference during repeated retrieval, regardless of whether the overlap is semantic or segmental, whether the modality of production is spoken or written, or whether the locus of selection is lexical nodes or segments. While the observed interference is believed to reflect a similar computational principle, the resulting effect is not necessarily expected to have identical properties at the two stages. For instance, the presence of intervening items seems to reduce the interference generated during lexical-segmental mapping more than it reduces the interference generated during semantic-lexical mapping, which typically survives lags of 10+ items (Schnur, 2014). Future work should investigate potential differences to more completely characterize the mechanisms at the two stages. However, this will not form a part of the current dissertation: instead, I will focus on the implications of these similarity effects for the (re)learning of words. At this stage, the data are equally consistent with an account that relies on lateral inhibition (Howard et al., 2006) and with an account that relies on competitive incremental learning (Oppenheim et al., 2010). The remainder of this dissertation will not attempt to compare these accounts directly but rather will investigate whether the predictions of incremental learning models hold when extended beyond production of known words to word learning.

Interference for both modalities of production. Second, the findings filled an important gap in research by extending the investigation of the mechanisms of lexical selection and segmental encoding to written word production. In the adult system, written words are produced over a much longer time-course and in a more serial fashion than are spoken words, which could influence the temporal dynamics of these systems. Despite such differences between the systems, a striking aspect of the results of the background studies was that they provided strong evidence of the similarity in activation and selection principles across written and spoken modalities. While the nature of segmental representations differs between the two dissociable modalities of spoken and written production (e.g., Shen et al., 2013), similar computational demands are involved in converting a semantic representation to a lexical node and then to its corresponding segments. This work was important not only for understanding the written production system itself, but also because it provided evidence of the domain-generality of the mechanism underlying the effects of similarity.

Both interference and facilitation exist within the word production system. Third, these results support the claim that the facilitation effects which have been reported for initial segment overlap both here and in previous reports arose at least partially outside the word production system since they disappeared when predictability was eliminated. However, there may also be facilitatory effects of similarity within the production system itself that are masked by the stronger interference effects. It is important to consider that facilitatory and inhibitory effects coexist in the word production system, and performance reflects the interaction of these opposing forces. Task affects this interplay such that semantic similarity typically creates interference when related pictures are named repeatedly (e.g., Damian et al., 2001), but it creates facilitation when a single semantically related word is presented (e.g., Wheeldon & Monsell, 1994). Similarly, phonological overlap can have both facilitatory and inhibitory effects depending on task. Even within the same task, facilitation and interference are present throughout and the shifting balance between them can be observed. In the blocked cyclic naming paradigm, there is facilitation due to repetition for all naming contexts that initially predominates (e.g., it is most evident in the drop in response times between cycles 1 and 2 in all four background experiments reported here) before interference is observed in later cycles due to presentation in contexts with high overlap. Note that this facilitation due to repetition is also due to long-term learning: strengthening of target connections makes it easier to produce the same word on later trials. The critical claim is that the long lasting learning that induces interference and facilitation is similar between semantic-lexical and lexical-segmental mapping.

Summary. In sum, the background work presented in this chapter provides evidence for the generality of the learning mechanisms that apply across spoken and written word production to both semantic and form-based levels of representation, leading to complex patterns of interference and facilitation. That is, the results aligned with the predictions of the incremental learning models of lexical selection proposed by Howard and colleagues (2006) and Oppenheim and colleagues (2010) and extended here to segmental encoding: there was evidence of interference resulting from both semantic and unpredictable segmental overlap. In later chapters of this dissertation, I will investigate the consequences that these mechanisms have for the learning of new words that must be incorporated into the extant production system, focusing on the impact of learning words in contexts of high similarity. In the next chapters, I will lay the foundations for this investigation by reviewing relevant literature regarding learning and develop a framework to extend an incremental learning model from production to learning.

CHAPTER 2: LEARNING AND DIFFICULTY

As established in the last chapter, producing words in the context of other related words (regardless of whether they are semantically or segmentally related) leads to interference compared to producing the same words in the context of unrelated words. Essentially, it is more difficult to say or write a word when one has previously produced related words: people are slower to produce the targets in such a context. What consequences does this have for learning new words? In this chapter, I will provide background to address this question, reviewing (1) the education literature regarding the benefits of introducing difficulty in learning environments; (2) the rehabilitation literature comparing errorless learning methods to those that require effortful processing and allow for errors; and (3) the second language learning literature regarding the effects of learning related words.

Desirable Difficulties

What is a Desirable Difficulty?

While one's immediate assumption may be that difficulty will have a negative impact on learning, there is in fact a large literature in the education field addressing the benefits of so-called "desirable difficulties." Desirable difficulties, a term coined by Robert A. Bjork and colleagues in the early 1990s, refers to the idea that a learning situation that requires effortful processing may initially decrease performance, but in fact may enhance long-term learning (e.g., E. L. Bjork, Bjork, & Mcdaniel, 2011; E. L. Bjork, deWinstanley, & Storm, 2007; R. A. Bjork & Bjork, 1992; R. A. Bjork, 1994, 2013; Karpicke, Lehman, & Aue, 2014; Karpicke & Roediger, 2007; Kornell & Bjork, 2007;

Roediger & Karpicke, 2006b). This idea depends on the distinction between learning and performance: one can perform well at a task in a particular context in the short term, but fail to learn in the long term or to apply knowledge to a new context. The converse is also possible: learning can occur even when there is not an immediate change in performance.

Bjork and colleagues have explained this distinction in terms of storage strength vs. retrieval strength (E. L. Bjork & Bjork, 2011; R. A. Bjork & Bjork, 1992). Storage strength refers to the quality of the memory representation (e.g., how associated it is with related knowledge and skills), while retrieval strength refers to the current accessibility of the representation. These two strengths are independent: some items have low storage strength but high retrieval strength (e.g., the name of the main meeting room at a conference hotel on the second day of the conference); some items have high storage strength but low retrieval strength (e.g., the address of the house you lived in for 5 years but have not thought about for 20 years); yet others have both high storage strength and high retrieval strength (e.g., the cell phone number you currently have and have had for over 10 years); while some have both low storage strength and low retrieval strength (e.g., the exact sentence someone said to you ten minutes ago). The conditions that most rapidly increase retrieval strength are not the same conditions that increase storage strength. In the conditions that rapidly increase retrieval strength, learners often misperceive the effects of study: since their performance improves rapidly, they believe they have learned the material (i.e., they conflate storage and retrieval strength). However, if storage strength is not increased, they will have poor performance after a delay or in a different context.

Despite their independence, storage strength and retrieval strength are related in that both are increased by effortful retrieval. According to this theory, retrieval is a powerful modifier of memory: when an item is retrieved from memory, that item's representation is strengthened such that it has greater storage strength and greater retrieval strength (R. A. Bjork, 1975). The amount of strengthening is a function of retrieval difficulty: when the retrieval process is more difficult (i.e., when the initial retrieval strength is lower), the increase in storage and retrieval strength is greater (R. A. Bjork & Bjork, 1992; R. A. Bjork, 1994). Therefore, introducing the right sort of difficulty—one that increases effortful retrieval—leads to increased long-term learning, compared to situations where no such desirable difficulty is involved.

The concept of desirable difficulties is not that *any* increase in difficulty will be beneficial (McDaniel & Butler, 2011). Rather, difficulties that allow for encoding and retrieval processes that enhance learning are helpful. If a person does not have the necessary background knowledge and skills to effectively process the difficulties, they are undesirable (E. L. Bjork & Bjork, 2011). For example, asking a second grader to complete a linear algebra problem will not help with their arithmetic skills even though linear algebra does rely on those skills to some extent. Since the child does not have the necessary background to successfully complete (or even begin) the task, it will constitute an undesirable difficulty. This is similar to the concept of the region of proximal learning, in which it is recommended that learners study materials that are not too difficult, but are just beyond what has already been mastered (Metcalfe, 2011).

It should be noted that the theory of desirable difficulties does not provide a mechanistic explanation for exactly how effortful retrieval increases a representation's storage strength or what makes a specific item difficult in a specific retrieval situation (Karpicke et al., 2014). However, the general principle can be successfully applied to explain results in a wide variety of situations, several of which are described below. That this theory does not offer specifics of the mechanism does not mean that it cannot be applied in the learning situation I am concerned with in this dissertation: contexts of high semantic and segmental similarity. In fact, as will be discussed later, the model of the production system described in the previous chapter provides a mechanism whereby the cause and effects of retrieval difficulty can be explained in this situation.

Examples of desirable difficulties. To provide some context for the way that desirable difficulties have been investigated previously, several examples are described.

Spacing. The first example of a desirable difficulty that has received attention is spacing, whereby distributed and massed practice are compared. Massed practice refers to a situation where study of a given item is not separated by time or by presentation of any other items. Distributed practice refers to the opposite situation where multiple study opportunities are separated by time and/or by intervening items (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Findings here reflect general knowledge about cramming: although students may be successful on a test if they intensively study shortly before taking it, they rapidly forget this information. If it is necessary to retrieve this information later (e.g., for a cumulative final exam), they have to study all over again. On the other hand, if students study less intensively over a longer time period, they are more likely to retain that information not only for the immediate test but also further into the future. Formal study of the spacing effect goes all the way back to Ebbinghaus (1885/1913), who found that he could learn to recite a 12-syllable series after 38 repetitions distributed over three days, but that he needed 68 immediately successive (massed) repetitions on the one day just previous to learn a similar 12-syllable series. He concluded that it is more advantageous to distribute repetitions over time than to mass them all at a single time point. Over the years, many researchers have directly investigated this phenomenon, finding consistent results that support distributed practice over massed practice (for review of the extensive literature, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006 and references within). Cepeda and colleagues (2006) conducted a meta-analysis of 839 assessments of distributed practice from 317 experiments with verbal memory tasks in 184 published articles. They found a robust advantage of distributed over massed practice regardless of the retention interval between study and test, which ranged from less than one minute to over thirty days. Cepeda and colleagues also analyzed the relationship between inter-study interval (time between practice sessions) and retention interval (time between practice and test), finding that the optimal inter-study interval was longer when the retention interval was longer. This suggests that to optimize long-term learning it is better not merely to have practice distributed over time as opposed to massed, but also to have larger lags between study sessions. Overall, distributed practice is consistently superior to massed practice. This spacing effect can be explained as a desirable difficulty. When there is more time or competing information between presentations of an item, retrieval of that item becomes more effortful. Attempting this effortful retrieval modifies the memory representation of that item, strengthening it to a greater extent relative to the situation where practice was massed and retrieval relatively easy.

Testing effect. Another example of a desirable difficulty is the testing effect, which refers to the finding that repeated testing leads to better long-term retention than repeated studying, even when no feedback is given about responses on the initial test (Hogan & Kintsch, 1971; McDaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006a; Wheeler, Ewers, & Buonanno, 2003; see Roediger & Karpicke, 2006b for review). This happens even though repeated studying leads to better performance on tests given immediately after learning. For example, Roediger and Karpicke (2006a) had students study prose passages. In a repeated testing condition, students then took one or three free-recall tests on the material without being given feedback on their responses. Alternatively, in a repeated study condition, students instead restudied the material one or three times. On an immediate final test five minutes later, those in the repeated study conditions did better than those in the repeated testing conditions. However, when the final test was delayed by two days or one week, those in the repeated testing condition demonstrated superior performance. According to a desirable difficulties explanation, those in a repeated testing condition have more practice with actually retrieving the items. Since the information is not as easily accessible as in a study condition (low retrieval strength), retrieval will be effortful and modify the memory representation by increasing storage strength. In restudying, retrieval of the information is easy since it is available to look at again. Here, the information has high accessibility (high retrieval strength), but not high storage strength. The retrieval that happens here will have less of an impact on storage strength than in the repeated testing condition. On an immediate test, repeated study leads to better performance because there is greater retrieval strength that has not yet decayed; but on a delayed test, repeated testing leads to better performance because there is greater storage strength established via effortful retrieval during learning that remains even though retrieval strength has decayed over time.

Errorless and Errorful Learning in Cognitive Rehabilitation

Benefits of Errors

Having established that there is a benefit of repeated testing, one may wonder if this benefit only exists when learners provide correct answers during the test or are given corrective feedback. This question has been investigated in the education literature, where it is found that testing is beneficial regardless of whether correct answers are given by the learner or as feedback (e.g., Roediger & Karpicke, 2006a), although there are additional benefits when feedback is given for incorrect responses (e.g., Pashler & Rohrer, 2007). For example, Kornell, Hays, and Bjork (2009) gave participants fictional general knowledge questions to which they could not possibly know the answer beforehand (e.g., Who shot a fig out of a tree with a crossbow in the 11th century?). Participants who were forced to guess before being told the answer performed just as well on a final cued recall test as subjects who initially read the answer along with the question for the same period of time, even though the answer was not available during the initial retrieval attempt (guess) for the first group. Similarly, when participants were asked to learn a pair of weak associates (e.g., pond-frog), those who were first asked to guess before being given the intended associate performed better on a final test than those who spent the same amount of time studying the complete pair. That is, generating an incorrect answer did not harm learning and in some cases in fact improved it, as long as the correct answer was given. Attempting to retrieve a piece of information, even one that is not known, benefits learning of that information. This could be because retrieval attempts result in deep processing at retrieval and deeper encoding when the answer is provided, because retrieval routes are strengthened, and/or because any information that is retrieved can later be used to cue the correct answer.

Errorless vs. Errorful Learning

This line of research is closely related to the debate in cognitive rehabilitation regarding errorless as compared to errorful learning (for review, see Middleton & Schwartz, 2012). The debate here centers on whether it is more beneficial for individuals with neurological damage who are (re)learning skills or knowledge to avoid making errors (*errorless learning*), for instance by being given a target response on every trial, or to attempt effortful retrieval that may result in an error (*errorful learning*). The idea behind errorless learning is that it avoids implicit learning of associations between a stimulus and an incorrect response that may be reinforced and produced in the future when an explicit error is made, while strengthening the association between the stimulus and correct response in line with the Hebbian learning principle that neurons that fire together wire together (Fillingham, Hodgson, Sage, & Lambon Ralph, 2003; Middleton, Schwartz, Rawson, & Garvey, 2015; Middleton & Schwartz, 2012).

Originally, errorless learning was studied in animals learning motor responses. For example, pigeons learned to discriminate red and green when an errorless learning method was applied (e.g., Terrace, 1963).

The debate was then extended to the cognitive domain, with applications to relearning of information by individuals with amnesia, who experience deficits in long-term, explicit, episodic memory (for review, see Clare & Jones, 2008). Here, there is generally a benefit of errorless learning (e.g., Baddeley & Wilson, 1994; Hunkin, Squires, Parkin, & Tidy, 1998; Page, Wilson, Shiel, Carter, & Norris, 2006; Squires, Hunkin, & Parkin, 1997). Note that there are exceptions depending on the severity of memory impairment and exact tasks and methods used, especially if memory impairment is mild to moderate and explicit recall is required. In such situations, the advantage for errorless learning disappears, and in a few cases a long-term advantage for errorful learning is revealed (e.g., Bier et al., 2008; Dunn & Clare, 2007; Evans et al., 2000; Haslam, Moss, & Hodder, 2010; Metzler-Baddeley & Snowden, 2005). One explanation for the general advantage of errorless learning is that individuals with severe amnesia, unlike neurotypical individuals, do not retain explicit memories of having made an error. They are therefore vulnerable to implicitly remembering the error that they produced instead of the correct response, even when given feedback, meaning that errors become self-reinforcing (for debate on this explanation, see Anderson & Craik, 2006; Baddeley & Wilson, 1994; Hunkin et al., 1998; Page et al., 2006; Tailby & Haslam, 2003). For such individuals, the typical long-term benefits of effortful retrieval (such as in the situations discussed above) do not outweigh the detriments of learning errors.

Errorless vs. errorful learning in aphasia treatment. What happens in aphasia? Since in this dissertation I am concerned with the learning of words, this population of people who experience difficulty with language due to an acquired brain injury is especially interesting. People with aphasia after stroke do not typically have the severe memory problems that lead to the advantage for errorless learning in amnesia. However, it is not uncommon for these individuals to have concomitant deficits in executive function and attention, which may impact their ability to detect errors and learn from feedback. This suggests that errorless learning methods may be more effective for this profile. On the other hand, those with primarily linguistic deficits may benefit from errorful learning's effortful retrieval in the same way as neurotypical individuals who, as reviewed above, do show increased learning when desirable difficulties are introduced. A number of studies have investigated whether errorless or errorful therapies are more beneficial for people with aphasia (for review, see Fillingham, Hodgson, Sage, & Lambon Ralph, 2003; Middleton & Schwartz, 2012), primarily concentrating on treatment of anomia, or word-finding difficulties.

Anomia treatment. Fillingham and colleagues (2003) presented a review of the anomia treatment literature from 1985-2003. While none of the studies included directly compared errorless and errorful methods, 72% of the 61 errorful interventions led to significant improvement immediately after therapy and 79% of the 29 errorless interventions did the same. At follow up, there was a numerical advantage for errorful therapies, with 59% of those studies demonstrating significant lasting effects while 47% of the errorless therapies showed similar significant effects. There was a similar trend for errorful therapies to result in generalization to control items more often (38% of studies) than errorless therapies (15% of studies). A secondary analysis looked only at expressive therapy types, the only type for which there were enough studies to compare errorless and errorful methods. Here, a greater proportion of errorless interventions were successful immediately after therapy (87% of errorless vs. 74% of errorful studies showed improvement), but this trend reversed at follow up (44% of errorless vs. 61% of errorful showed lasting effects) and on generalization measures (44% of errorless vs. 58% of errorful showed improvement on untreated items). These results suggest that both methods are effective, especially when evaluated directly after therapy, but that errorful treatments for anomia may have some advantages in the long term.

Since then, several studies have directly compared errorless and errorful treatments for anomia (Conroy, Sage, & Lambon Ralph, 2009; Fillingham, Sage, & Lambon Ralph, 2005a, 2005b, 2006; McKissock & Ward, 2007; Middleton et al., 2015). In the first of set of closely-related studies (published out of order), Fillingham, Sage, and Lambon Ralph (2006) gave a series of people with aphasia both an errorless therapy, in which participants named pictures that appeared with both spoken and written names, and an errorful therapy, in which participants attempted to name the pictures and then were given progressive phonemic and orthographic cues if their responses were incorrect. They found that both therapies were effective for most participants when assessed immediately post-treatment: eight participants showed a significant effect of both therapies, one of whom improved more with errorful therapy; one showed significant improvement for errorless therapy but not for errorful therapy, although the outcomes of two therapies did not differ significantly; and two showed no significant effect of either therapy. When long-term learning was assessed five weeks after therapy, there were indications that errorful learning was superior: four showed long-term benefits only for errorful therapy, two of whom showed a significant difference between errorful and errorless methods, while for the remaining participants there were no differences between errorful and errorless learning. Furthermore, Fillingham and colleagues (2006) found that therapy outcomes were correlated with non-language cognitive scores on a battery of assessments of recognition memory, executive function, and monitoring. These cognitive scores were also correlated with the advantage for errorful over errorless learning, indicating that individuals with the best memory and executive function were most likely to benefit from therapies incorporating effortful retrieval (see also Lambon Ralph, Snell, Fillingham, Conroy, & Sage, 2010 for evidence that these cognitive factors predict response to errorful treatment along with severity of language impairment). This is in line with explanations of errorful learning that suggest the benefits rely on error detection and monitoring abilities.

Two follow up studies by the same authors reach similar conclusions. Even when feedback is removed, errorful learning is as effective as errorless learning immediately after training (Fillingham et al., 2005a, 2005b). Comparing these three studies provides additional evidence that retrieval practice can enhance learning. In Fillingham and colleagues (2005a), the errorful therapy involved participants making three naming attempts for each picture, which was presented with the first phoneme and grapheme. They learned a greater proportion of words than in Fillingham and colleagues (2005b), a study in which they made only one attempt when trained with this therapy. In fact, the proportion of words learned when making three attempts was very similar to the first study, Fillingham and colleagues (2006), which included up to three attempts per picture but also had feedback. Based on these results, it seems that it is retrieval attempts themselves that enhance learning, while direct feedback does not confer an additional benefit. Most people with aphasia were able to learn just as much in errorful treatment conditions when attempting to provide names about which they were never given feedback as they learned in errorless treatment conditions when expressly given the names, indicating that this type of difficulty can be beneficial and enhance learning. Similar results were found when treatments for verbs were considered alongside nouns

(Conroy et al., 2009). The authors did note that participants preferred the errorless treatment conditions because they were less frustrating (Fillingham et al., 2006). However, the important point is that they were able to learn under errorful conditions, and that there was some evidence of a long-term advantage for this type of treatment.

Other investigations have found generally consistent results: errorful treatments tend to be just as effective as errorless. However, recent studies have emphasized the role of successful retrieval in errorful treatments. In contrast to errorless learning, a successful retrieval attempt during errorful learning is a correct response that is self-generated as the result of effortful retrieval, not one that is provided. Successful retrieval attempts are contrasted with unsuccessful retrieval attempts from the same errorful paradigm. Findings show that retrieval that leads to a correct response has a more beneficial effect than retrieval that leads to an incorrect response. For instance, Middleton and colleagues (2015) found that errorless treatment was less beneficial than errorful treatment, regardless of whether or not the errorful treatment included cuing, on a delayed test one day after training, but found that this advantage only persisted for the errorful treatment including cuing one week later. These authors directly examined the role of successful naming trials by modeling correct retrieval (successful errorful learning), incorrect retrieval (failed errorful learning), and repetition (errorless learning) trials during training. At follow up, items that had been retrieved correctly during training were more likely to be correct, whereas items that had been retrieved incorrectly did not significantly differ from the items that had merely been repeated during training. This suggests that both errorless and errorful therapies can be effective, but that maximizing successful retrieval can confer a learning advantage. However, when retrieval is unsuccessful, feedback may add an additional benefit. For example, McKissock and Ward (2007) found that errorful treatment with feedback led to increased word learning as compared to errorful treatment without feedback. These results differ from previous studies in which Fillingham and colleagues reported no additional advantage of feedback. This may be because the previous studies involved cuing that may have increased successful retrieval events as in Middleton and colleagues (2015).

Dysgraphia treatment. The exploration of errorless and errorful learning has also been extended to treatments of acquired dysgraphia. Raymer, Strobel, Prokup, Thomason, and Reff (2010) compared an errorless spelling treatment in which individuals with spelling deficits copied words while given progressively fewer cues to an errorful treatment in which they spelled words to dictation while given progressively more cues as necessary. They found that both methods produced positive results: all four of the participants in their study learned the words and one month later showed increased accuracy relative to pre-training for both errorless and errorful treatments. There was some advantage for errorful training in that three of the four participants had larger training and maintenance effect sizes for errorful training, although one showed the reverse pattern. It should be noted that, as in anomia treatment studies, participants preferred the errorless training because it was less frustrating, even though errorful training was more effective for most of them. The results of this study are consistent with the results of the anomia treatment studies discussed above: while both errorless and errorful treatments are successful, there is some evidence that errorful treatments may have an advantage especially in terms of long-term retention. Further support for the effectiveness of both errorless and errorful training for spelling impairments comes from a study by Thiel and Conroy (2014). These authors report that three of their four participants with dysgraphia made gains that were maintained five weeks after treatment for both errorless training consisting of copying and errorful training consisting of spelling to dictation accompanied by progressive orthographic cues. For one of these participants, the treatment effect was larger for the errorless than the errorful method. Other cognitive deficits prevented an additional participant from benefitting from either therapy. Together, these dysgraphia treatment studies provide evidence that there are benefits of both errorless and errorful training when the modality of treatment is extended from spoken to written production.

Summary. Overall, the evidence across spoken and written treatment studies is clear: people with aphasia improve after both errorless and errorful training in both the spoken and written modalities of word production. In general, the individuals tend to prefer errorless learning because it is less frustrating, but there may be some long-term retention advantages for errorful learning, especially when this method leads to successful retrieval attempts. This literature is consistent with the notion that introducing desirable difficulties is a valid strategy to help people with aphasia relearn words. Later in this dissertation, I describe an intervention for dysgraphia that I implemented based on this principle.

Desirable Difficulties Based on Context: Interleaved and Blocked Presentation

Interleaving as a Desirable Difficulty

Having established that introducing desirable difficulties that promote effortful retrieval can have beneficial effects on learning for both neurologically healthy adults and people with aphasia, I next consider whether introducing difficulty via manipulations of the context of items learned together constitutes a desirable difficulty. Up to this point, the difficulties considered were applicable regardless of the content being learned: they primarily related to the timing of training (e.g., spacing effects[9]) and the presence or absence of retrieval in the method of training (e.g., testing effect, errorless and errorful therapies). Here, the focus shifts to the composition of the materials that are to be learned. The effects of this type of difficulty are especially important here as one of the goals of this dissertation is to investigate what effect similarity among items trained together has on learning of those items. Specifically, I will investigate the effects of blocking training trials by semantic or segmental similarity as compared to training unrelated items.

Motor skills. In the literature concerning desirable difficulties, there are some indications that interleaving information has beneficial effects when compared to blocking information (E. L. Bjork & Bjork, 2011). This difference is especially clear for learning motor skills. One illustration of this comes from Shea and Morgan (1979), who trained healthy undergraduate participants to perform three motor tasks that involved responding to stimulus lights by knocking over movable barriers in different, specific orders. Note that this study did not require retrieval: a diagram of which barriers were to be hit in which order was provided on every trial. Instead, it examined the speed with which participants completed the tasks. Training trials were either blocked by task or presented in a random/mixed order. They found that those participants who were trained on the task in a mixed order were slower during acquisition of the skills than those

[9] Note that manipulating time does manipulate content: introducing spacing necessarily causes interleaving in that other information will be presented between presentations of the content at issue. However, much of the discussion of spacing centers on manipulations of time that do not depend on the amount or type of content present between study or test trials (e.g., separating study sessions by more or less time but not controlling what students do during that interval).

trained in a blocked order, but that those trained on a mixed order performed better on retention testing, completing tasks faster after a 10-day interval. This retention difference was especially pronounced when participants were tested in a mixed order. Here, those who had been trained in a blocked order performed especially slowly compared to those who had been trained in a mixed order. Similar results were found when participants were tested in a blocked order. In this situation, those trained in a mixed order were slightly faster than those trained in a blocked order. This study demonstrates that interleaving information initially increases difficulty during acquisition, but leads to better long-term outcomes.

Cognitive skills: Mathematics. Evidence that interleaving can be a desirable difficulty is not limited to motor skill learning. There is also evidence of interleaving being beneficial in cognitive domains. For example, Rohrer and Taylor (2007) examined the effect of blocking mathematics practice problems. They taught students the formulas to calculate the volumes of four geometric solids in two training sessions separated by one week. In each session, students in the blocked condition were given a tutorial on one solid followed by four practice problems related to it, a process that was repeated for each of the four solids. Students in the mixed condition were given all four tutorials followed by sixteen randomly ordered practice problems. The same tutorials and practice problems were used in both conditions, although critically in different orders. During practice sessions, those in the blocked condition were significantly more likely to be correct than those in the mixed condition. However, when students were tested one week after the final training session, the pattern reversed: now those trained in the mixed condition outperformed those trained in the blocked condition. This provides further evidence that interleaving trials has long-term benefits over blocking them.

Cognitive skills: Induction. The superiority of training with interleaved as opposed to blocked practice also extends to information acquired through induction (i.e., learning of principles, patterns, and concepts from observation of exemplars). Kornell and Bjork (2008) directly tested this by training participants to recognize artists' styles. Training materials consisted of six paintings by each of twelve artists. In the blocked condition, participants saw all six paintings by the same artist together. In the mixed condition, paintings by different artists were interleaved. Learning was assessed by showing 48 new paintings by the same artists and having subjects identify which artist had painted each. While participants who completed both conditions believed that blocked presentation was more helpful, results showed that performance was significantly better for artists whose styles were learned in the mixed condition. At the individual level, 78% of participants did better with mixed presentation than with blocked. A similar advantage of interleaving was found when the final test consisted of recognizing new paintings as by a familiar or unfamiliar artist instead of identifying the specific artist. The authors attribute the benefit of this difficulty to its role in enhancing discrimination. Juxtaposing the paintings of different artists may have highlighted the relative differences in style, helping participants learn to make appropriate discriminations. Discrimination is an important aspect of learning in many domains, including word learning, and situations that enhance the ability to make such discriminations are important learning opportunities. If interleaving in fact enhances discrimination abilities, it is a powerful tool. Regardless of whether this is the correct explanation, interleaving does seem to be a desirable difficulty in some learning domains, including motor skills, mathematics, and induction of patterns like artists' styles.

Interleaved and Blocked Presentation in Word Learning

Notice that none of the studies on interleaving were directly concerned with the learning of words. While interleaving practice on motor skills, mathematics, and inductive learning of categories may be beneficial, the situation may be different for words. The goals of these tasks are different: in word learning, individuals must learn a distinct, arbitrary word form for each object, while in mathematics and category learning, individuals must learn to apply the same non-arbitrary general principles (e.g., a formula or style) to different situations. Earlier in this dissertation I presented evidence from neurotypical participants demonstrating that producing the known names of pictures is in fact more difficult when those pictures are presented in semantically or segmentally similar blocks as opposed to unrelated. While overall participants are no less accurate in blocked contexts, they are slower. This difficulty may additionally be present during learning of new words that have to be integrated into the productions system, as opposed to solely after words have been fully learned. The situation where blocking makes production of known words more difficult contrasts with the examples of interleaving presented above where blocking initially made performing skills easier. It is possible that this increased difficulty due to blocking for words may enhance learning. According to this hypothesis, interleaving might be detrimental to the learning of words as opposed to beneficial as it has been previously described for learning of other types of content.

Part of this potential difference may depend on variations in what counts as blocked or mixed in these different situations. In the previously described literature, blocked items primarily are repetitions of the exact same stimulus or involve exactly the same concept. In the blocked naming that is of interest here, blocking refers to presentation of similar words, not the same ones. In fact, under the blocked cyclic naming paradigm utilized in the background experiments, there is interleaving of the same item in both blocked and unrelated conditions when words are repeated across cycles. Returning to the explanations of desirable difficulty suggested in the education literature, retrieval effort may vary in the different situations. When blocked items consist of immediate repetitions of the same stimulus or concept, retrieval is simple (retrieval strength is strong), and thus the representation's storage strength is not strongly modified. On the other hand, when blocked items consist of similar items, retrieval is more effortful: not only are there multiple items to remember, but they must also be discriminated between so that the correct concept is selected. This effortful retrieval may lead to greater strengthening of the representation. Blocking by similarity may have different effects than blocking by the same item or concept.

Blocking by Semantic Category

There have been a few investigations of the effects of similarity on word learning. The most direct investigations of this are rooted in the literature on second language learning. As with learning new words, learning words in a second language requires creating new links between meanings and forms. In this domain, the debate on whether words from the same semantic category should be taught together remains open.[10] It is

[10] Note that in this dissertation, when I discuss semantic relatedness I refer to categorical relationships. The role of associative relationships will not be considered. Associative and categorical relationships have been shown to have very different effects on production by neurotypical and aphasic participants, as well as on learning (e.g., Crutch, Connell, & Warrington, 2009; Crutch, Ridha, & Warrington, 2006; Crutch & Warrington, 2005, 2007, 2010; Higa, 1963; Nation, 2000; Tinkham, 1997).

very common for textbooks to be organized by semantic category, with chapters focusing on vocabulary related to topics like food, travel, family members, and school. Some have argued that this organization by category is beneficial because it reflects the organization of the lexicon (in terms of the architecture presented here, this includes the connections between the semantic features and lexical nodes), much of which learners already have through their first language (Channell, 1981; Erten & Tekin, 2008; Hashemi & Gowdaseaei, 2005)[11]. However, the observation that there is categorical organization in the language system does not mean that presenting items from the same category together is optimal for learners acquiring vocabulary. While such presentation could help learners by encouraging deeper levels of processing and allow items in sets to reinforce one another (Craik & Lockhart, 1972; Hashemi & Gowdaseaei, 2005), that is not necessarily the case. Indeed, many empirical studies argue that presenting items in semantic sets leads to interference (Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Higa, 1963; McGeoch & McDonald, 1931; Papathanasiou, 2009; Schneider, Healy, & Bourne, 1998; Tinkham, 1993, 1997; Underwood, Ekstrand, & Keppel, 1965).

While the impact of associative relationships on learning is certainly an interesting question that should be examined in future studies, it is beyond the scope of the current dissertation.

[11] Categorical organization of lexical-semantic information is widely supported by other psycholinguistic findings. There is neuropsychological evidence that there are distinct semantic categories that can be differentially impaired by brain damage (e.g., Capitani, Laiacona, Mahon, & Caramazza, 2003; Caramazza & Mahon, 2003, 2006; Caramazza & Shelton, 1998; Laiacona, Barbarotto, & Capitani, 2006; Mahon & Caramazza, 2009; McCarthy & Warrington, 1994; Shallice & Warrington, 1984; Tyler & Moss, 2001; Warrington & McCarthy, 1994, 1983), as well as neuroimaging evidence that there are brain regions that respond differentially to distinct semantic categories (e.g., Binder, Desai, Graves, & Conant, 2009; Cappa, Perani, Schnur, Tettamanti, & Fazio, 1998; Chao, Haxby, & Martin, 1999; Chao, Weisberg, & Martin, 2002; Chao & Martin, 2000; Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004; Devlin, Moore, et al., 2002; Devlin, Russell, et al., 2002; Gerlach, 2007; Goldberg, Perfetti, & Schneider, 2006; Grossman et al., 2002; Moore & Price, 1999; Mummery, Patterson, Hodges, & Wise, 1996; Perani et al., 1999; Phillips, Noppeney, Humphreys, & Price, 2002; Pilgrim, Fadili, Fletcher, & Tyler, 2002). Furthermore, short-term memory studies in which neurotypical participants recall or recognize lists of items typically show facilitation when the lists are categorized, especially if participants are given category labels during study (e.g., Bousfield, 1953; Bower, Clark, Lesgold, & Winzenz, 1969; Cole, Frankel, & Sharp, 1971; D'Agostino, 1969; Gollin & Sharps, 1988; Moely & Shapiro, 1971; Pollio, Richards, & Lucas, 1969; Sharps, Wilson-Leff, & Price, 1995; Toglia, Hinman, Dayton, & Catalano, 1997; Tulving & Pearlstone, 1966; Tversky, 1973).

Much of this evidence for interference as a result of presenting semantically related items is drawn from studies in which learners were taught new, artificial labels for L1 words. For example, Finkbeiner and Nicol (2003) had English-speaking undergraduate participants learn pseudoword names for common pictures. They found that immediately after training, participants were faster to translate items that had been learned in unrelated contexts in which items from different categories were mixed as opposed to those that were learned in blocks of items drawn from the same category. That is, they were slower to retrieve new words that had been trained in a related context, suggesting that semantic blocking increases difficulty. Other studies have also investigated whether semantic blocking is harmful or helpful by using the time or number of trials necessary to learn the labels as the dependent measure. Tinkham (1993) reported that English-speaking adults were faster to learn word pairs consisting of pseudowords coupled with unrelated English words than they were to learn word pairs consisting pseudowords coupled with semantically related English words in that they required fewer trials to reach criterion performance immediately after training. Tinkham (1997) and Waring (1997) reported similar findings when training was extended to the written modality (i.e., participants wrote responses after reading stimuli instead of saying responses aloud after hearing stimuli) and to a different language (Japanese): participants consistently required more trials to reach criterion on the semantically related pairs than the unrelated pairs. Taken together, these studies demonstrate that semantic blocking leads to interference during learning.

When the depth and breadth of vocabulary knowledge is probed instead of the time to reach criterion or immediate performance, results look different with regard to the advantages and disadvantages of semantic blocking. For example, Hashemi and Gowdasiaei (2005) observed adults in Iran learning real English words that were either blocked by semantic category or unrelated. Participants did learn in both conditions, but those who were trained with words blocked by category made greater gains in vocabulary knowledge as assessed by self-report immediately and one week after training than those who were trained with unrelated words. Another classroom study, this time of English-speaking undergraduates learning Japanese, demonstrated increased retention of words learned from a list of categorically related items as compared to unrelated lists or lists related in different ways when participants were tested 3-4 days after receiving lists to be studied at home (Hoshino, 2010). Other evidence tentatively supporting the idea that the difficulty of learning words in semantic blocks may improve outcomes comes from Erten and Tekin (2008), who studied Turkish fourth graders learning English words that were either blocked by category (e.g., food, animals) or unrelated. As reported in previous studies, when students were tested immediately after training, they were more accurate on the words that had been trained in unrelated blocks. However, on delayed tests the situation was less clear. While overall there was still an advantage for the unrelated words, this advantage was carried by the most recently presented sets that had been learned one week before. When the sets that had been learned two weeks before were examined, there was no significant difference between related and unrelated and, in fact, students were numerically more accurate on the words from the related set. These results suggest that, while there is initially interference for semantically related words that are being learned, as time progresses this disadvantage is reduced and may be reversed in the long term. Such a pattern is in keeping with the idea that semantic blocking may be present a desirable difficulty during learning: while it initially reduces performance, it may lead to better long-term retention.

A few studies do provide evidence against this view. For instance, Papathanasiou (2009) trained Greek-speaking learners of English on words that were semantically blocked or unrelated in a classroom setting. Immediately after training, adult beginning learners performed significantly better on words trained in the unrelated context than on words trained with semantically related words, in line with the previous studies suggesting that semantic blocking increases difficulty in a way that disadvantages words during and immediately after training. However, this study found that the disadvantage persisted: the difference was still statistically significant when tests were given again two weeks later. As a caveat, note that only three of the six semantic blocks used in training consisted of categorically related words, while the other three were made up of homonym, synonym, and antonym pairs. This makes it less clear that the difference is truly about categorical semantic relatedness, the type under investigation in this dissertation.

Another study presenting contradictory results is Schneider, Healy, and Bourne (2002), who argue that interleaving is a desirable difficulty in word learning. English-speaking undergraduates learning French vocabulary made fewer errors during an initial learning session when words were blocked by category as opposed to unrelated, but made more errors during a relearning session one week later. Based on this data, the authors argue that there is an immediate advantage for words learned in blocks, indicating greater contextual interference in unrelated blocks, but the effect reverses over time, suggesting that the interference enhances long-term learning. This argument runs counter to the evidence I have presented suggesting exactly the opposite, that blocking by semantic category, not interleaving, is the desirable difficulty in word learning. It is not immediately apparent why these authors observed results that contradict many of the studies presented above, in which there was an immediate benefit of training words in unrelated blocks and sometimes a long-term advantage of training in semantically related blocks. It should be noted that they did not find any differences between words trained in blocked and unrelated contexts on a retention task at the beginning of the second session one week after the first, which does raise some questions about the strength of their results and interpretation. This study does agree with the important point that the conditions that most benefit acquisition are not necessarily those that most benefit retention. Further investigation of whether semantic blocking is a desirable difficulty that creates such conditions that benefit long-term learning is necessary.

Blocking by Segmental Similarity

To date, there has been less investigation of the effects of learning or seeing vocabulary in segmentally related sets. However, the evidence available suggests that this type of blocking leads to interference (Nation, 2000). For example, Talamas, Kroll, and Dufour (1999) found that English-Spanish bilinguals experienced interference during a translation-verification task when given false translation pairs in which the given incorrect translation was related in form to the correct translation (e.g., garlic-ojo [eye] instead of garlic-ajo). This interference was especially pronounced for those who were less fluent in the second language. For these participants, this interference as a result of segmental relatedness was greater than the interference observed for false translation pairs in which the given translation was semantically related to the correct translation (e.g., garlic-cebolla [onion] instead of garlic-ajo). Altarriba and Mathis (1997) reported similar results: monolingual English speakers who were taught Spanish words were slower to respond on a translation recognition task when presented with orthographically related foils as compared to unrelated foils. That it is the less fluent bilinguals as well as the beginning second language learners who experience the most interference for segmentally related pairs suggests that this type of similarity may cause interference effects during learning, especially at early stages. Note, however, that this study does not measure interference during acquisition of those words, but after they have been at least partially learned.

In a different study, Creel, Aslin, and Tanenhaus (2006) had participants learn pseudowords for novel objects, half of which were phonologically similar to one other included item. Across testing sessions immediately after training on three consecutive days, participants were more likely to error by matching a heard word to a distractor picture that was phonologically similar to the target in either its rhyme or onset than to an unrelated distractor picture. This provides evidence that phonologically similar words trained in the same session interfere with one another, starting early in learning.

Another line of research that indirectly addresses the question of whether segmental similarity is helpful for learning concerns the effects of neighbors. Neighbors are commonly defined as words that differ from each other by one phoneme (phonological neighbors) or one grapheme (orthographic neighbors) (Coltheart, Davelaar, Jonassen, & Besner, 1977). A word with many neighbors is said to have a dense neighborhood, while a word with few neighbors is said to have a sparse neighborhood (Luce & Pisoni, 1998). Increased phonological neighborhood density generally leads to interference in the spoken production of known words, although conflicting results have been observed (for review, see Chen & Mirman, 2012; Sadat, Martin, Costa, & Alario, 2014). On the other hand, there is some evidence that orthographic neighborhood density facilitates written production (Roux & Bonin, 2009). What effect does neighborhood density have on learning? Research on this topic is relevant to investigation of the role of segmental similarity between materials trained together because it addresses the question of how newly learned items are integrated with other words that have similar forms. In studies of neighborhood density, these similar words are not explicitly presented, but they are activated along with the item being learned and constitute an important part of the learning environment. One might expect similar effects when the context of segmentally similar words is actively manipulated by presenting neighbors together during training.

One example of a study manipulating neighborhood density of words to be learned is Storkel, Armbrüster, and Hogan (2006), in which adults learned pseudoword names for novel pictures presented in story contexts. On an immediate test requiring production of the names of the objects given the pictures, participants were more accurate on pseudoword names with higher neighborhood density regardless of whether the pseudowords had high or low phonotactic probability. Here, having more real word neighbors facilitated learning, possibly by allowing for faster integration into the lexicon as connections are built to known related words.

Stamer and Vitevitch (2012) similarly reported that increased neighborhood density facilitates learning of new words. In this study, increased neighborhood density facilitated the learning of Spanish words by English speakers. Participants performed more accurately on tasks involving picture naming and picture-word matching when the newly learned word sounded similar to many known Spanish words as compared to few. This advantage for high neighborhood density was observed both immediately and 48-72 hours after training.

A further study by Storkel, Bontempo, and Pak (2014) extended these results. As in previous studies, they found that there were immediate advantages of high neighborhood density for the learning of novel words paired with novel pictures as measured by a picture naming task. Furthermore, they reported that when a second set of novel words was trained, participants were more successful with novel words that were phonological neighbors of those in the first set. This suggests there may be an immediate advantage for words that are similar in form. However, the effects reversed directions when performance was tested again one week later. In this situation, there was more forgetting for the high density words and those that sounded similar to recently named words. This result suggests that low phonological neighborhood density may be a desirable difficulty. Based on this, one might infer that dissimilar words should be treated together. At first, this result seems to conflict with studies reported above that find interference effects early in learning when form related words are presented together. One possibility is that there may be important differences in the consequences of similarity when it is actually present in the materials trained together as opposed to when it is implicitly present in the production system into which new items are being integrated. Alternatively, differences may be due to different time points after training being considered as early and late learning. Some of the early learning studies showing interference may be more comparable to the later time point in this study and be a better reflection of long-term learning. If so, it might truly be that segmental similarity leads to interference when long-term learning is assessed, supporting the interpretation that low segmental similarity may be a desirable difficulty.

On the other hand, there is evidence in the neuropsychological literature that training neighbors can have a positive impact on the relearning of word spellings. Sage and Ellis (2006) reported that training a set of words led to improved spelling of the trained set's untrained orthographic neighbors for a dysgraphic individual with a graphemic buffer deficit. In a related vein, Raymer, Cudworth, and Haley (2003) reported that a different individual with a graphemic buffer deficit had improved spelling that generalized to untrained words that shared orthography with the trained words. Kohnen, Nickels, Coltheart, & Brunsdon (2008) reported a similar finding in a child who experienced dysgraphia after a traumatic brain injury. Here, orthographic neighborhood size was a good predictor of which untrained words improved after treatment, and the untrained words most likely to benefit were orthographically similar to the trained words. These studies suggest that the relationship between words with similar written forms can be beneficial for learning: the shared activation of orthographically related words does not lead to insurmountable interference and in fact can lead to improved outcomes for a particular type of patient with a particular type of deficit. Although the related words are not presented together, relearning one spelling can have an implicit positive impact on its neighbors. It remains to be seen if this positive impact remains when the relationship between the neighbors is emphasized by explicitly presenting the words in the same context. Note that it has been previously reported that damage to the production system can change the balance of facilitation and interference observed in neighborhood density experiments, so these results should be interpreted cautiously (Sadat et al., 2014).

Overall, the impact of segmental similarity on learning should be investigated further, with the goal of determining whether increasing or reducing the degree of form similarity of items trained together is beneficial. From the current evidence, it seems that form dissimilarity (low segmental similarity) may be a desirable difficulty. High segmental similarity during learning seems to lead to interference when that learning is later assessed, but may have facilitatory effects during learning. However, reports are not entirely consistent. Some studies suggest that high segmental similarity leads to interference during training as well. Others show that high segmental similarity may be beneficial in treatment for some individuals with dysgraphia. Segmental similarity is clearly a factor that requires further investigation.

Summary

This chapter has presented evidence that there are desirable difficulties that can enhance learning. That is, manipulations that increase retrieval difficulty or depth of encoding may initially have a negative effect on performance during or immediately after training, but ultimately have a positive effect when long-term learning is considered. A few of these desirable difficulties include distributed as opposed to massed practice and multiple tests as opposed to repeated study opportunities. Introducing such difficulties in training can benefit not only neurotypical individuals, but can also aid individuals with aphasia, who must relearn language knowledge and skills after deficits caused by brain injury. This is shown through studies of errorless and errorful learning, which generally find that errorful treatments have at least as good of outcomes as errorless treatments and may confer some long-term advantages.

In this dissertation, I am particularly interested in the effects of context similarity on learning or relearning words (i.e., training segmentally or semantically related words together as opposed to training unrelated words together). This raises the question of whether similarity between words is a desirable difficulty that should be introduced to improve learning outcomes. Past research suggests that in several domains, including motor and mathematical skill acquisition, interleaving information is a desirable difficulty. This would suggest that words should be presented in unrelated contexts. However, when word learning is specifically considered and blocks include repetitions of similar words as opposed to repetitions of the exact same concept, the situation changes. Here, past research suggests that training items in blocks arranged by semantic category initially leads to interference, but there is some evidence of positive effects on long-term learning, meaning that semantic similarity is likely a desirable difficulty. There is less direct investigation of training with segmentally related blocks. Some evidence suggests that, in contrast, form *dissimilarity* may be the desirable difficulty with high form similarity leading to relative facilitation during training and interference at later assessment, although other reports support the idea that high form similarity causes interference during training as well. Contrasts between the effects of semantic and segmental blocking may arise because representations are strengthened differently by retrieval in the various contexts. In the next chapter, I will consider explanations of these potential desirable difficulties based on the extension of the model of production presented in the first chapter of this dissertation. These lead to testable predictions about how similarity will impact both neurotypical individuals learning novel words for novel items and people with aphasia relearning previously known written spellings.

CHAPTER 3: A FRAMEWORK FOR UNDERSTANDING THE EFFECTS OF BLOCKING

In the previous two chapters, I have described an incremental learning account of written and spoken production that is supported by background experiments, and I have reviewed the literature on the impact of difficulty on learning, including effects of similarity blocking. In this chapter, I bring these lines of research together to present a framework for understanding the consequences of similarity in the learning environment. This framework extends the incremental learning models from production of known words to the learning of new words and generates testable predictions about the impact of semantic and segmental similarity during training on long-term learning outcomes. In the remainder of the dissertation, I will test these predictions using experiments with neurotypical participants learning new words and individuals with dysgraphia relearning spellings of previously known words. Note that this is confirmatory science investigating whether these predictions hold, not a direct comparison of alternative accounts since at present there is not a clear alternative to incremental learning accounts and since the implemented incremental learning models proposed by Howard and colleagues (2006) and Oppenheim and colleagues (2010) make similar predictions.

Learning: Balancing Similarity and Distinctiveness.

In order to make predictions about conditions that optimize learning, it is first helpful to consider what is involved in the word learning task. Essentially, a learner needs to create a new association between a semantic representation, a lexical representation, and the segments used to produce that lexical representation. For simplicity, consider the task of learning a novel name for a novel concrete object when explicitly given the pairing between a picture of the object and the name (thus avoiding the problem of determining the referent, an important yet complicated component of learning from naturalistic input). To successfully learn the name,[12] a learner must extract semantic features from the picture, connect these to a new lexical node based on those features, and connect appropriate phonemes or graphemes of to that new lexical node so that the name can be produced. This new item must be integrated into the mental lexicon that already exists along with previously known words as well as with novel items that are being learned at the same time. Because of this integration, there will be activation not only of the word being learned, but also of other items that share semantic features and that overlap in segments. This activation of non-target items can lead to interference (increased difficulty in production). An important part of the learner's task is to develop a distinctive representation for the new word that differentiates it from the representations of similar words. Developing distinctiveness may be aided by comparison to similar words, which can provide context that contrasts fine-grained differences of the words. The task of learning a new word critically involves balancing the demands of avoiding interference from similar words while enhancing distinctions between the new word and closely related words (see discussions of interference theory and distinctiveness theories in the L2 acquisition literature, e.g., Nation, 2000; Papathanasiou, 2009; Tinkham, 1997; Waring, 1997).

[12] In the present discussion, I conceptualize representation and process separately (e.g., semantic features represent content that is connected to the separate content of lexical nodes through the process of lexical selection). This is not to say that such distinctions are necessary in every framework. Some connectionist approaches postulate distributed representations that do not make the distinction between representation and process. Instead, in these models, information is encoded in the weights of connections between nodes: there is no information stored in the nodes themselves without the weighted connections.

An Example: Learning Non-Native Speech Contrasts

One example of balancing similarity and distinctiveness comes from the literature on adults learning to perceive speech contrasts that are not used in their native language. A common example of this is Japanese-speaking adults learning to perceive the distinction between the phonemes [r] and [l] in English. In Japanese, these sounds are perceived as the same phoneme, and Japanese learners of English often have persistent difficulty perceiving and producing the contrast that can be difficult to remediate, even after years of exposure to English (e.g., Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Bradlow, 2008; Iverson, Hazan, & Bannister, 2005; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991). One series of analyses incorporating computational simulations and empirical investigation suggests that in order for Japanese speakers to learn to make the discrimination, they need to be trained in a way that makes the differences between [r] and [l] apparent to them (McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; McClelland, Thomas, McCandliss, & Fiez, 1999). Initially, these learners cannot hear the difference between the phonemes in naturalistic speech: they are mapped onto the same phoneme category based on previous experience with Japanese, and typical input simply reinforces these categories (McClelland et al., 1999). Empirically, Japanese-speaking learners of English who are trained using natural stimuli of [r] and [l] without any feedback do not learn to distinguish between them (McCandliss et al., 2002). Given the right environmental conditions, however, the difference can be learned. One way to accomplish this is to train the learners using highly exaggerated stimuli. Under these conditions, the difference can be perceived and new categories begin to form. Over time, the amount of exaggeration can be gradually reduced, and the new categories can be reinforced regardless of whether or not feedback is given. Eventually, learners can perceive the distinction in natural stimuli. A different but similarly successful method is to train using natural stimuli that learners cannot initially distinguish, but to give feedback as learners practice making the distinction. This can help them to direct attention to the relevant differences and to create consistently labeled representations of the two phonemes as training proceeds (McCandliss et al., 2002). Both of these training methods led learners to successfully distinguish between [r] and [l] after training. What the successful training methods have in common is that they help learners attend to the distinctions, which allows them to overcome interference due to similarity. Balancing these demands of similarity and distinctiveness can lead to positive learning outcomes.

Beneficial Conditions for Learning

The goals of word learning thus include not only creation of the connections necessary for lexical access and segmental encoding of the new word, but also integration into the production system that balances similarity and distinctiveness in order to maximize long-term learning outcomes. When considering the potential advantages and disadvantages of training words in the context of similar vs. unrelated words, it is important to consider these goals. One must recall that the conditions that are best during acquisition are not necessarily the best for long-term learning. Difficulty during learning (e.g., creating interference by presenting similar words together) may turn out to be beneficial if it leads to stronger long-term representations. One way in which long-term representations could be stronger is if they are more distinct (i.e., if the contributions of distinctive, unshared features are enhanced while the contributions of shared features are reduced) so that it is easier to distinguish the item from related items. Keep in mind that not every type of difficulty will improve learning outcomes. For instance, if learners cannot perceive the differences necessary to make distinctions between items, they will not form distinct representations. In such a case, the interference caused by presenting similar items would lead to negative effects. In this situation, the long-term representations might be weaker and less distinctive, with increased contributions of shared features but not of distinctive features (so that it is harder to distinguish the item from related items). Overall, training in blocks of related items may be beneficial if it increases the distinctiveness of representations as compared to training in unrelated blocks, but it will be detrimental if it instead decreases distinctiveness.

Extended Incremental Learning Model Predicts the Effects of Blocking on Word Learning

The incremental learning account of the production system presented earlier in this dissertation makes different predictions about whether blocking will increase distinctiveness based on the type of similarity. In the following example, I will consider the predictions of the incremental learning model of lexical selection and the extension of it to segmental encoding that I have proposed[13]. I will assume that this model also describes the processing that occurs when new words are learned since the same type of

[13] The incremental learning model discussed here is based on the Oppenheim and colleagues (2010) model. I do not use the Howard and colleagues (2006) model here because it does not make clear predictions about differences in shared and distinctive semantic features. However, the predictions that it makes regarding segmental similarity are in line with those of the extended Oppenheim and colleagues (2010) model. Because the Oppenheim and colleagues (2010) model and its extension do make clear predictions about blocking's effects on distinctiveness based on the type of similarity and, furthermore, because they align with the predictions of the Howard and colleagues (2006) model regarding segmental similarity, I focus on it for clarity.

connections need to be formed between semantic features of the new item and its lexical node and since new words need to be integrated into the production system so that they can be used. This extension of the incremental learning model of production to learning is a novel contribution of this dissertation.

Predictions Regarding Semantic Blocking

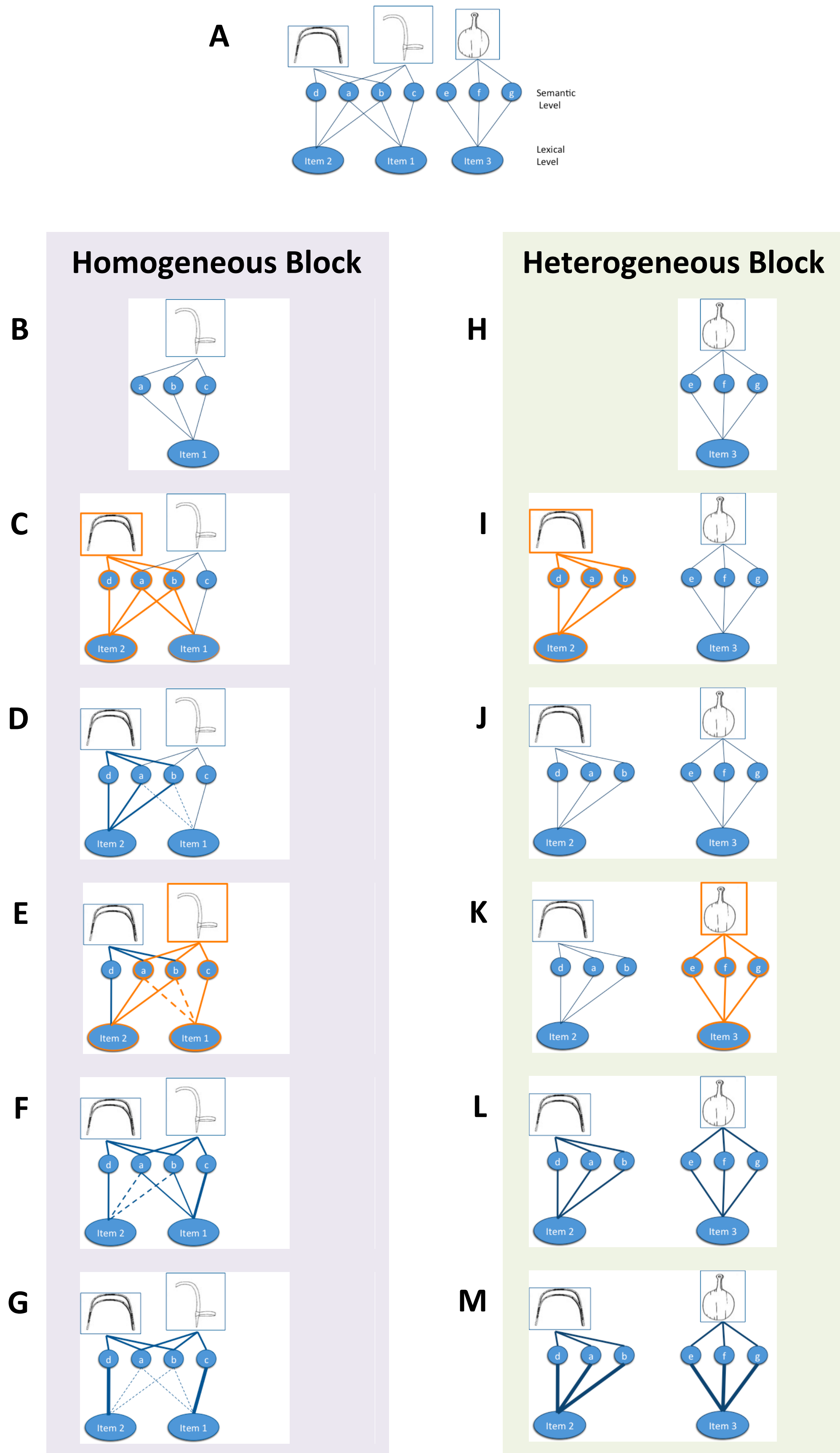
As reviewed in the first chapter, according to the Oppenheim and colleagues (2010) incremental learning model, lexical selection is a competitive learning process: connections between semantic features and lexical nodes are strengthened if they contribute to activation of the selected lexical node but weakened if they contribute to activation of a different lexical node. Interference, as observed in decreased speed and/or accuracy, arises when semantically related words are presented together because the connections between shared features and lexical nodes are weakened. That is, after the first item is selected, connections between its semantic features and lexical node are strengthened because they contributed to activation of the target. Connections between semantic features that the target shares with other items and the lexical nodes of those items are weakened because they contribute to activation of non-targets. When one of these related items is subsequently presented as a target, its lexical node will receive less activation from those shared semantic features because the connections were weakened on an earlier trial. It will thus be more difficult to produce than an unrelated item whose connections were not modified by previous production since the unrelated item does not share semantic features with the previously produced word.

What happens when this model is extended from production of known words to the learning of new words? I illustrate my proposed interpretation of the extension with an example (Figure 10)[14].

Figure 10. Example implementation of an incremental learning model of lexical selection extended from production models to address learning of new items.

Panel A shows the network that the learner is acquiring. There are three pictures, with connections to seven semantic features and then to three lexical nodes. Two of the lexical nodes, Item 1 and Item 2, are semantically related, each sharing two of their three semantic features, while Item 3 is not semantically related to the other items. Panels B-G, in the purple box, depict learning in a semantically homogeneous context. Here, the semantically related Item 1 and Item 2 are learned. Panels H-M, in the green box, depict learning in a semantically heterogeneous block. Here, the semantically unrelated Item 2 and Item 3 are named. Orange lines depict activation during each trial. Blue lines depict the connections between levels. The weight of these lines is used to depict strengthening and weakening, with thicker, solid lines showing strengthened connections, and thinner, dashed lines showing weakened connections. Refer to the text for a detailed description of the example depicted in this figure.

[14] This model (as well as the one that will be described when discussing predictions about segmental blocking) is obviously simplified: it focuses on the words being learned without taking into account the other words in the system that the learner already knows. Within the set of known words there are likely many items that are related in meaning and/or form. These relationships to previously known words may impact the learning of new words, leading to complex interactions. One limitation of the work presented here is that it does not consider these relationships. In the experiments that follow, I assume that the words being learned together are most prominent and likely to have the largest impact on one another throughout the experiments. This assumption may be incorrect: the larger set of words with which the newly learned items are integrated may greatly impact the learning process.


A
Semantic Level
Lexical Level
Homogeneous Block
Heterogeneous Block
B
C
D
E
F
G
H
I
J
K
L
M
d
a
b
c
e
f
g
Item 1
Item 2
Item 3

Panel A shows the network that the learner is acquiring. There are three pictures, with connections to seven semantic features and then to three lexical nodes. Two of the lexical nodes, Items 1 and 2, are semantically related. Here, a and b are shared features since both lexical nodes are connected to these features, while c and d are distinctive features since only Item 1 is connected to semantic feature c and only Item 2 is connected to semantic feature d. Item 3 is not semantically related to the others. This lexical node is connected to semantic features e, f, and g, which are all distinctive features since they do not activate any other lexical nodes in this network. Note that all three items are novel objects with unknown names for the learner at the beginning of training: this network structure does not yet exist in the learner's mind.

Panels B-G illustrate what happens in a semantically homogeneous learning context.

In panel B, a picture of the first target, Item 1, is shown. In this learning environment, semantic features a, b, and c send activation to the new lexical node of Item 1. When it is selected, the connections between all three semantic features and Item 1's lexical node are strengthened. At this point, Item 1 is the only item being learned, so no other representations are active and thus there is no other strengthening or weakening.

In panel C, Item 2 is introduced. In this learning environment, semantic features a, b, and d send activation to the new lexical node of Item 2. Because Item 1 was learned on the previous trial, shared semantic features a and b also send activation to its lexical node. Item 2 is more active than Item 1 because it receives activation from more semantic features (feature d in addition to a and b), and its lexical node is selected when it reaches the selection threshold.

Panel D shows that when Item 2's lexical node is correctly selected, its connections to semantic features a, b, and d are strengthened because they contributed to correct activation of the target. Meanwhile, connections between Item 1's lexical node and semantic features a and b are weakened because they contributed to activation of a non-target. Note that it is the connections between shared semantic features and lexical nodes that are weakened, while the connection between the distinctive semantic feature and its lexical node are strengthened.

Panel E shows a second practice trial with Item 1. As in panel B, semantic features a, b, and c send activation to the lexical node of Item 1. Features a and b also send activation to the lexical node of Item 2, as in panel C. Since the connections between the lexical node of Item 1 and semantic features a and b were weakened on the previous trial when Item 2 was practiced, semantic features a and b now send less activation to Item 1's lexical node than they would have sent before practicing the related item. Thus there is interference: selection of Item 1 is slower after having produced Item 2.

Panel F shows that, after Item 1 is correctly selected, the connections between its lexical node and semantic features a, b, and c are strengthened because they contribute to activation of the target. Meanwhile, the connections between Item 2's lexical node and semantic features a and b are weakened because they contributed to activation of a non-target. Note that it is again the connections between shared features and lexical nodes that are weakened, while the connection between the distinctive feature and its lexical node are strengthened.

Panel G shows what happens over time as the semantically related items are practiced repeatedly. In this situation, the connections between shared features and lexical nodes are repeatedly weakened because they contribute to the activation of non-targets. However, the connections between the distinctive features and lexical nodes are always strengthened because they only contribute to activation of correct targets. Therefore, distinctiveness of the representations is enhanced.

Contrast panels B-G with panels H-M, which illustrate what happens in a semantically heterogeneous learning context.

In panel H, Item 3 is introduced. In this situation, it is the first item to be introduced. In the learning environment, semantic features e, f, and g send activation to the new lexical node of Item 3. When it is selected, the connections between all three features and Item 3's lexical node are strengthened. At this point, Item 3 is the only item being learned, so no other representations are active and thus there is no other strengthening or weakening.

In panel I, Item 2 is introduced. In this learning environment, semantic features a, b, and d send activation to the new lexical node of Item 2. Item 3 is not related to Item 2, and these are the only items in the system, so the active semantic features only send activation to the lexical node of Item 2.

Panel J shows that when Item 2's lexical node is correctly selected, the connections between its lexical node and semantic features are strengthened because they contributed to correct target activation. The connections between Item 3's lexical node and semantic features are not involved, and do not undergo strengthening or weakening.

Panel K shows a second practice trial with Item 3. As in panel H, semantic features e, f, and g send activation to the lexical node of Item 3. Since Items 2 and 3 do not share features, there is no activation of Item 3. The connections between Item 3's lexical node and semantic features were never weakened. Therefore, in contrast to panel E, there is no interference.

Panel L shows that after Item 3 is correctly selected, the connections between its lexical node and features a, b, and c are strengthened because they contribute to activation of the target. The connections between Item 2's lexical node and semantic features are not involved, and do not undergo strengthening or weakening.

Panel M shows what happens over time as the unrelated items are practiced over and over. In this situation, the connections between all features and the lexical nodes are repeatedly strengthened. There is no weakening of connections since none of the semantic features contribute to activation of non-target lexical nodes. The distinctiveness of the representations is unchanged throughout learning.

Overall, this model predicts that semantic blocking is a desirable difficulty relative to training in unrelated blocks: while there may be interference during acquisition, this interference increases the distinctiveness of the representations trained in semantically related blocks. The connections between distinctive semantic features and lexical nodes will be increased while the connections between shared semantic features and lexical nodes will be reduced. If increased distinctiveness enhances long-term learning outcomes as expected, training in semantic blocks should be beneficial in the long run. This model predicts that the representations of words trained in semantically related blocks will benefit from becoming more distinctive and thus easier to retain in long-term memory as compared to words trained in unrelated contexts.

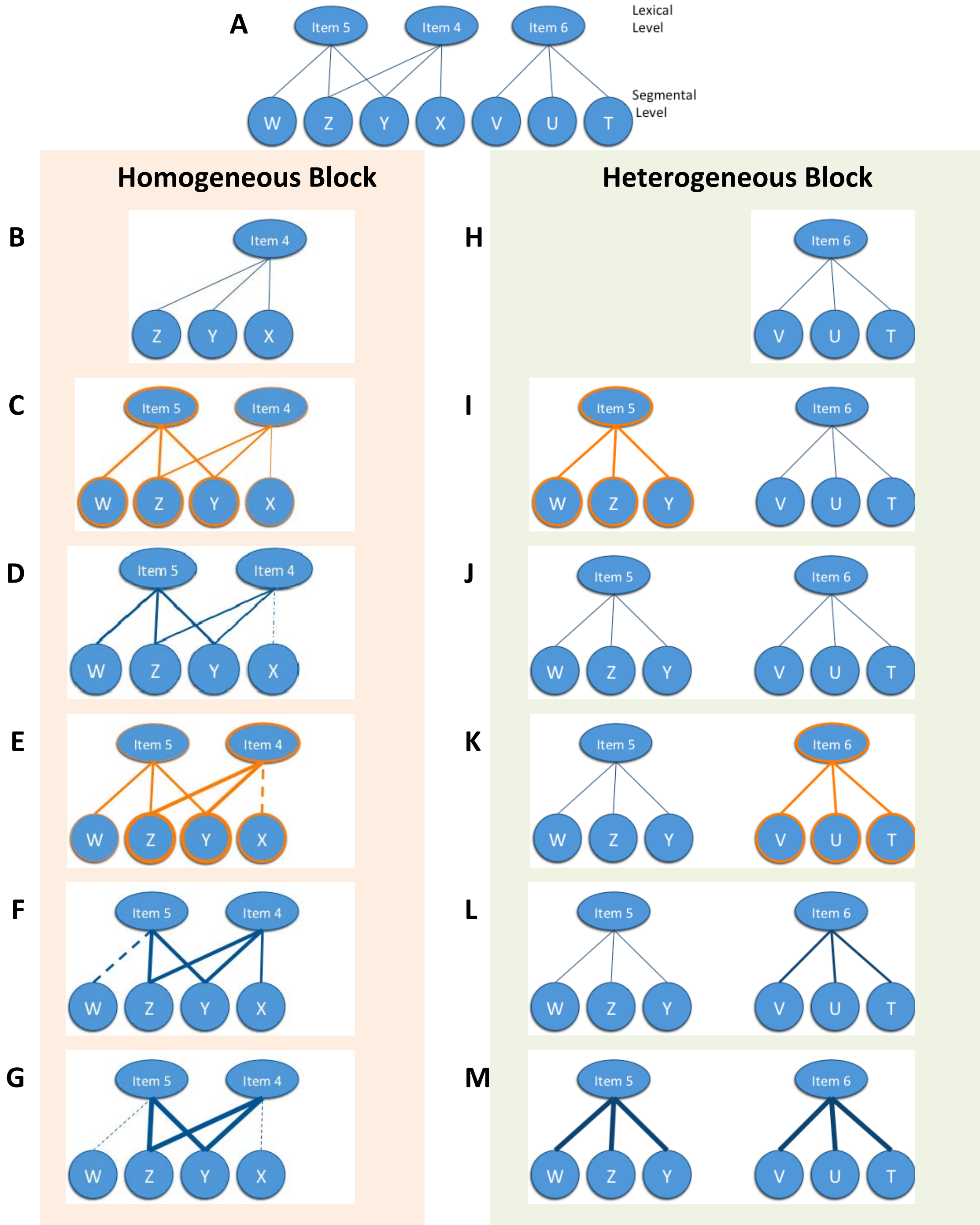
Predictions Regarding Segmental Blocking

I now turn to the extension of the incremental learning model to segmental encoding presented in Chapter 1 of this dissertation. The same general principles apply as in lexical selection: there is strengthening of connections that contribute to activation of the correct target, but weakening of connections that contribute to the activation of non-targets. Additionally, under the interpretation presented earlier, there is feedback between segments and lexical nodes, which leads to activation of lexical nodes that share segments. According to this model, interference arises when segmentally related words are presented together. In contrast to the situation in lexical encoding, this is not because the connections to shared features are weakened, but because the connections to distinctive segments are weakened. When the first target is selected, the connections between target segments (both shared and distinctive) and both target and non-target lexical nodes that were activated through feedback are strengthened, while the connections between the non-target lexical nodes and non-target segments (i.e., the distinctive segments that are not shared with the target) are weakened. When the next target is a related item that shares segments with the previous target, it is more difficult to produce. This is because the connections between its lexical node and its distinctive segments were weakened on the previous trial because they were active but did not contribute to correct production of the target. This difficulty is relative to the situation where the two targets are unrelated, meaning there is no shared activation and thus no weakening of connections between the lexical node and segments of the second target when the first is produced.

Next, I extend this production model to learning. An example of this is depicted in Figure 11.

Figure 11. Example implementation of an incremental learning model of segmental encoding extended from production models to address learning of new items.

Panel A shows the network that the learner is acquiring. There are lexical nodes with connections to seven segments. Two of the lexical nodes, Item 4 and Item 5, are segmentally related, each sharing two of their three segments, while Item 6 is not semantically related to the other items. Panels B-G, in the orange box, depict learning in a segmentally homogeneous context. Here, the segmentally related Item 4 and Item 5 are learned. Panels H-M, in the green box, depict learning in a segmentally heterogeneous block. Here, the segmentally unrelated Item 6 and Item 5 are named. Orange lines depict activation during each trial. Blue lines depict the connections between levels. The weight of these lines is used to depict strengthening and weakening, with thicker, solid lines showing strengthened connections, and thinner, dashed lines showing weakened connections. Refer to the text for a detailed description of the example depicted in this figure.


A
Item 5
Item 4
Item 6
Lexical Level
Segmental Level
W
Z
Y
X
V
U
T
Homogeneous Block
Heterogeneous Block
B
C
D
E
F
G
H
I
J
K
L
M

Panel A shows the network that the learner is acquiring. There are three lexical nodes, with connections to seven segments. Two of the lexical nodes, Items 4 and 5 are segmentally related. Here, Z and Y are shared segments since both lexical nodes are connected to these segments, while X and W are distinctive features since only Item 4 is connected to segment X and only Item 5 is connected to semantic feature W. Item 6 is not segmentally related to the others. This lexical node is connected to segments V, U, and T, which are all distinctive since they are not connected to any other lexical nodes in this network. Note that the learner knows none of the three items at the beginning of training: this network structure does not yet exist in the learner's mind.

Panels B-G illustrate what happens in a segmentally related learning context.

In panel B, Item 4 is introduced. In this learning environment, Item 4's new lexical node sends activation to segments Z, Y, and X. Because Item 4 is the only item in the network at this point, feedback from the segments does not lead to activation of any additional lexical nodes. When the segments are selected, the connections between all three segments and Item 4's lexical node are strengthened. Since no other representations are active, there is no other strengthening or weakening.

In panel C, Item 5 is introduced. In this learning environment, Item 5's new lexical node sends activation to segments W, Z, and Y. Because Item 4 was learned on the previous trial, shared segments Z and Y also send activation to its lexical node via feedback. Item 4's lexical node in turn sends activation to its segments, including segment X that is not part of target Item 5. Overall, Item 5 is more active than Item 4 and its segments are selected when the selection threshold is reached.

Panel D shows that when Item 5's segments W, Z, and Y are correctly selected, their connections to the lexical node of Item 5 are strengthened because they contributed to correct activation. Connections between the selected segments Z and Y and the lexical node of Item 4 are also strengthened because they too contributed to correct activation of the target, Item 5. Meanwhile, the connection between the active, non-target segment X and Item 4's lexical node that contributed to its activation is weakened. Note that here it is the connection between the distinctive non-target segment and its lexical node that is weakened, while the connections between shared target segments and the lexical nodes (both the target and non-target but related lexical nodes) are strengthened.

Panel E shows a second practice trial with Item 4. As in panel B, the lexical node of Item 4 sends activation to segments Z, Y, and X. Through feedback, segments Z and Y send activation to the lexical node of Item 5, which in turn sends activation to its segments, including W that is not part of target Item 4. Since the connections between Z, Y, and both lexical nodes were strengthened on the previous trial, there is now more activation of both lexical nodes that there would otherwise have been. Furthermore, the connection between distinctive segment X and the lexical node of Item 4 was weakened on the previous trial, so that segment receives less activation than it would have received before practicing the related item. Thus there is interference: selecting the segments for Item 4 is slower after having produced Item 5.

Panel F shows that, after Item 4's segments is correctly selected, the connections between those segments and the lexical nodes of Items 4 and 5 are strengthened because they contribute to correct activation of the target. Meanwhile, the connection between Item 5's lexical node and distinctive segment W is weakened because it contributed to activation of a non-target segment. Note that it is again the connection between the distinctive segment and its lexical node that is weakened, while the connection between the shared segments and both lexical nodes are strengthened.

Panel G shows what happens over time as the segmentally related items are practiced repeatedly. In this situation, the connections between shared segments and lexical nodes are always strengthened because they only contribute to the activation of target segments. However, the connections between the distinctive features and lexical nodes are repeatedly weakened because they contribute to activation of non-target segments. Therefore, the distinctiveness of the representations is reduced.

Contrast panels B-G with panels H-M, which illustrate what happens in a segmentally unrelated learning context.

In panel H, Item 6 is introduced. In this situation, it is the first item to be introduced. In the learning environment, the new lexical node of Item 6 sends activation to segments V, U, and T. When they are selected, the connections between all three segments and Item 6's lexical node are strengthened. Since Item 6 is the only item in the network at this point, there is no feedback to related lexical nodes and no other strengthening or weakening.

In panel I, Item 5 is introduced. In this learning environment, Item 5's new lexical node sends activation to segments W, Z, and Y. Item 6 does not share segments with Item 5, and these are the only items in the system, so there is no activation of Item 6.

Panel J shows that after Item 5 has successfully been selected, the connections between its lexical node and segments are strengthened because they contributed to correct target activation. The connections between Item 6's lexical node and semantic features are not involved, and do not undergo strengthening or weakening.

Panel K shows a second practice trial with Item 6. As in panel H, the lexical node of Item 6 sends activation to segments W, Z, and Y. Since Items 5 and 6 do not share segments, there is no activation of Item 5. The connections between Item 6's lexical node and segments were never weakened. Therefore, in contrast to panel E, there is no interference.

Panel L shows that after Item 6 is correctly selected, the connections between its lexical node segments are strengthened because they contribute to activation of the target. The connections between Item 5's lexical node and segments are not involved, and do not undergo strengthening or weakening.

Panel M shows what happens over time as the unrelated items are practiced over and over. In this situation, the connections between the lexical nodes and segments are repeatedly strengthened. There is no weakening of connections. The distinctiveness of the representations is unchanged throughout learning.

Overall, this model predicts that segmental blocking is not a desirable difficulty relative to training in unrelated blocks. As with semantic blocking, segmental blocking is expected to lead to interference during acquisition. In contrast to semantic blocking, this interference reduces the distinctiveness of the representations trained in segmental blocks: the connections between lexical nodes and distinctive segments are reduced, while the connections between lexical nodes and shared segments are increased. While participants must still create distinct representations in order to learn the words, doing so will be more difficult in a context of segmental blocking as compared to an unrelated context since the strengthening and weakening of connections leads to relatively less distinctive representations. If reduced distinctiveness leads to worse long-term learning outcomes as expected, training in segmental blocks should be detrimental in the long run. This model predicts that the representations of words trained in segmentally related blocks will become less distinctive and thus harder to learn and retain as compared to words trained in unrelated contexts.

Summary of Predictions Regarding Blocking

This chapter presented an extension of the incremental learning model of production to learning (from this point forward, referred to as the Extended Incremental Learning Model and abbreviated e-ILM). The e-ILM generates testable predictions regarding the effects of blocking during learning on the trajectory and outcomes of learning based on changes to the distinctiveness of representations. That is, this type of model claims that there are different consequences of semantic and segmental blocking on shared and distinctive features or segments and relates these changes to theories of long-term learning and retention. For each type of blocking, there are four predictions, summarized in Table 4.

Table 4. Predictions of the e-ILM regarding the effects of training in semantic and segmental blocks.

Semantic Blocking	***Segmental Blocking***
1. Training in semantic blocks leads to interference during acquisition relative to training in unrelated contexts.	1. Training in segmental blocks leads to interference during acquisition relative to training in an unrelated context.
2. Training in semantic blocks increases the distinctiveness of representations by strengthening the connections between distinctive semantic features and lexical nodes while weakening the connections between shared semantic features and lexical nodes.	2. Training in segmental blocks reduces the distinctiveness of representations by strengthening the connections between lexical nodes and shared segments while weakening the connections between lexical nodes and distinctive segments.
3. Training in semantic blocks is beneficial for retention relative to training in unrelated contexts, assuming that increased distinctiveness leads to better long-term learning outcomes.	3. Training in segmental blocks is detrimental for retention relative to training in unrelated contexts, assuming that reduced distinctiveness leads to worse long-term learning outcomes.
4. Semantic blocking is a desirable difficulty: increased training difficulty (interference) leads to increased long-term retention due to increased distinctiveness.	4. Segmental blocking is not a desirable difficulty: increased training difficulty (interference) leads to reduced long-term retention due to reduced distinctiveness.

The first two predictions for each type of blocking are direct consequences of extending the incremental learning models from production of known words to the learning of new words. They deal with the relative strengthening and weakening of connections as a result of training in blocked vs. unrelated contexts. While both semantic and segmental blocking should create interference during acquisition relative to learning of unrelated items, this interference is due to different strengthening and weakening of shared vs. distinctive features. Training in semantic blocks strengthens the connections between distinctive semantic features and lexical nodes while weakening the connections between shared semantic features and lexical nodes. This means that training in semantic blocks should increase distinctiveness. On the other hand, training in segmental blocks strengthens the connections between lexical nodes and shared segments while weakening the connections between lexical nodes and distinctive segments. This means that training in segmental blocks should reduce distinctiveness.

The second two predictions rely on assumptions that are made to relate these models that address what occurs during training to effects on long-term learning. The third predictions rely on the assumption that long-term learning outcomes are better for items that have more distinctive representations. If this is the case, training in semantic blocks should lead to better retention than training in unrelated contexts since it enhances distinctiveness, but training in segmental blocks should lead to worse retention than training in unrelated contexts since it reduces distinctiveness. The fourth predictions relate these models to the hypothesis of desirable difficulties. Instead of predicting that any situation that increases difficulty during training will increase long-term learning, the fourth predictions takes into account the previous predictions to describe the situations in which blocking constitutes a desirable difficulty and when it does not. According to the e-ILM, semantic blocking is a desirable difficulty: it increases interference during training, enhances distinctiveness, and therefore should also increase long-term retention relative to training in unrelated contexts. Segmental blocking is not a desirable difficulty: it also increases interference during training, but it reduces distinctiveness, and so should decrease long-term retention relative to training in unrelated contexts.

There are also several alternative predictions that should be considered. (1) It could be that the interference during acquisition caused by blocking of both types is insurmountable and will lead to negative long-term learning outcomes regardless of how distinctiveness is changed. (2) Another possibility is that it is simply difficulty that matters for long-term learning outcomes, not the underlying cause of the difficulty. In this case, in line with the general hypothesis of desirable difficulties, the interference caused by both semantic and segmental blocking during acquisition might enhance long-term learning outcomes. (3) Of course, it is also possible that one or more of the specific e-ILM predictions are wrong in that training in blocked contexts does not lead to interference during training, that distinctiveness is not changed as expected, or that enhancing distinctiveness does not enhance long-term learning as assumed.

In the following two chapters, I describe two studies that directly tested the predictions of the e-ILM. The first study (Chapter 4) investigated neurotypical participants learning pseudoword names for novel objects. Across participants, I examined the effects of learning new words in semantically related, segmentally related, and unrelated blocks. Beyond examining how well participants acquired the words after training in different conditions by testing production immediately after training and at a later time point, I also probed the distinctiveness of the representations to see if different training contexts led to differences in speed or accuracy in responding to shared vs. distinctive semantic features and segments. The second study (Chapter 5) investigated a practical application of blocking. Here, I examined the relearning of previously known written word forms for objects by people with acquired dysgraphia. In this experiment, I compared the effects of training the spellings of previously known words in semantically related, orthographically related, and unrelated blocks. This allows for testing of the incremental learning model in a different, clinical population. Together these studies provide an opportunity to investigate the predictions of the e-ILM and enhance the theoretical understanding of the language learning system, but they may also have important educational and clinical implications for the training of words in practice.

CHAPTER 4: STUDY 1. LEARNING OF NEW WORDS IN NEUROTYPICAL ADULTS

Introduction

The goal of this study was to test the e-ILM predictions presented in the last chapter. In this study, neurotypical participants were taught novel names for pictures of novel objects, as well as receiving information about the semantic features of each item. Different groups of participants were trained on spoken and written production. Items were trained in blocks that were semantically related, segmentally related, or unrelated. Learning was assessed both immediately after training and several weeks later. In addition to naming accuracy and speed, the learning of distinctive vs. shared features was directly examined during each session. The goal of this study was to shed light on the consequences of different types of blocking, allowing examination of whether each type of blocking has beneficial or detrimental effects both during training and in the long run. Furthermore, the study investigated the interpretation of these effects that attributes them to differential weighting of distinctive and shared features. If the e-ILM applies, then semantic blocking should decrease the strength of shared features while increasing the strength of distinctive features. This shift means that the representations trained in this context should be more distinctive than those of items trained in the context of unrelated items. Increased distinctiveness should provide long-term benefits for learning outcomes even while there is interference during acquisition. The same type of model also predicts that segmental blocking should increase the strength of shared features while decreasing the relative strength of distinctive features. The resulting representations should be less distinctive than those of items trained in the context of unrelated items, which should lead to negative learning outcomes in the long term in addition to interference during acquisition. Additionally, including training in both modalities of production allows investigation of whether the effects of blocking are replicated across modalities, potentially providing further support for similar underlying mechanisms of spoken and written production.

This study used a training paradigm loosely related to the Ancient Farming Equipment paradigm (e.g., Cornelissen et al., 2004; Grönholm, Rinne, Vorobyev, & Laine, 2005; Grönholm-Nyman, Rinne, & Laine, 2010; Hultén, Vihla, Laine, & Salmelin, 2009; Laine & Salmelin, 2010; Tuomiranta, Grönholm-Nyman, Kohen, Martin, & Laine, 2011; Tuomiranta, Rautakoski, Rinne, Martin, & Laine, 2012). This paradigm, which has been used to investigate the neural correlates as well as the behavioral effects of new word learning, involves training participants to name ancient Finnish farming equipment. The stimuli are pictures of real objects, although participants are extremely unlikely to recognize any of them. The real Finnish names for the objects have been used for Finnish participants. For them, these are unknown words that preserve the phonotactic and orthotactic structure of Finnish: they are essentially pseudowords (Tuomiranta et al., 2012). For English speaking participants, modifications have been made to the names to make them phonotactically plausible English pseudowords (Tuomiranta et al., 2011). Although various modifications of training have been applied, the most common procedure used by this group is computerized training over four sessions in which each object is presented with its name (and, in some conditions, its definition) four times per session (Grönholm-Nyman et al., 2010; Grönholm et al., 2005; Tuomiranta et al., 2011, 2012). Participants are asked to read the name (and definition) aloud, or to listen to the name (and definition) when each picture is presented and then to repeat the name aloud. During each session and at follow-up, memory for the names of the pictures is tested via confrontation naming. Results show that this procedure leads to effective learning of both name and definition (Grönholm-Nyman et al., 2010). Participants are typically well over 90% accurate by the fourth session on both names and definitions. Although there are many reports of successful training of novel pairings of artificial, foreign, or uncommon words and pictures of familiar or novel objects in both neurotypical and cognitively impaired participants (e.g., Basso, Marangolo, Piras, & Galluzzi, 2001; Creel et al., 2006; Finkbeiner & Nicol, 2003; Freed, Marshall, & Phillips, 1998; Gupta, Martin, Abbs, Schwartz, & Lipinski, 2006; James & Gauthier, 2004; Kapnoula, Packard, Gupta, & McMurray, 2015; Kelly & Armstrong, 2009; Lotto & de Groot, 1998; Mayberry, Sage, Ehsan, & Lambon Ralph, 2011; Storkel et al., 2006, 2014; Tinkham, 1997; Waring, 1997), this paradigm seems particularly relevant for the present study because definitions are taught as well as items and because the method has been applied to treatment of anomia (e.g., Tuomiranta et al., 2011, 2012). Furthermore, this training is similar to that applied in Study 2, a dysgraphia treatment study that is presented in the next chapter. In the present study examining new word learning in neurotypical participants, I used a similar protocol to train items in the spoken version of the experiment while adapting the procedures for the written version of the experiment. The set of novel items was expanded to include a wider variety of objects beyond ancient Finnish farming equipment. Instead of sentential definitions, a list of four semantic features was trained along with each name. Semantically related, segmentally related, and unrelated blocks of different items were trained in the same participants, and the effects of this blocking on learning trajectories and outcomes were examined.

In addition to testing production of the new words after training to see how the training context influences learning outcomes both immediately and at follow up, this study also evaluated effects of training on the distinctiveness of the representations through the use of probe tasks. To compare shared and distinctive features at the semantic level, a semantic feature probe task was administered. Also known as a feature verification task, this task has been used previously to investigate the structure and organization of semantic representations (e.g., Cree, McNorgan, & McRae, 2006; McRae, Cree, Westmacott, & de Sa, 1999; Randall, Moss, Rodd, Greer, & Tyler, 2004; Rips, Shoben, & Smith, 1973). In the task, participants are presented with both a picture and a feature and asked to indicate whether or not the feature is true of the depicted concept. In the present study, the relationship between the features and items that yield "yes" responses was manipulated. When words are trained in a semantically similar context, a feature can be shared with the other category members that were also trained in that block, or a feature can be distinctive to that item. If training in a semantically related context enhances distinctiveness as predicted by the incremental learning model, participants should respond more slowly to shared features compared to distinctive features trained in semantic blocks and compared to features for words that are not trained in semantically overlapping contexts (which are not shared within a block). This prediction is in line with previous results suggesting that individuals are faster to respond to distinctive features than shared features in a feature verification task when the stimuli are well known words (Cree et al., 2006). Successful training in the semantically related condition may enhance this pattern.

To compare distinctive and shared features at the segment level, a letter probe task was used for participants in the written version of the experiment. Similar letter probe tasks, also called spelling probe tasks, have been used previously to evaluate the activation of orthographic representations (Rapp & Kong, 2002; Rapp & Lipka, 2011). In the letter probe task, participants see a picture from the training set along with a printed letter. Their task is to respond “yes” by pressing one button if the printed letter is in the spelling of the picture’s name or “no” by pressing a different button if it is not. In this study, the relationship between letters and items that yield “yes” responses was manipulated. When words are trained in a segmentally similar context, the letter in question might be shared with other items that were trained in the same block, or it might be distinctive to that item. If training in a segmentally related context decreases distinctiveness as is hypothesized, participants should respond more quickly to shared letters compared to distinctive letters of words trained in segmental blocks and compared to letters of words that are not trained in segmentally overlapping contexts (which are not shared within a block). In the spoken version of the experiment, an analogous sound probe task was used that pairs a picture and a recorded phoneme, asking the participant to indicate whether or not the given sound is in the name of the picture.

The study presented here was designed to test the predictions of the e-ILM. It examined whether training in semantically related or segmentally related blocks relative to unrelated blocks increased training difficulty, impacted the distinctiveness of the presentations being learned, and affected long-term learning outcomes. Together, the results of these analyses also allow for evaluation of whether each type of blocking is a desirable difficulty.

Method

Participants

Thirty-four neurotypical adult participants aged 18-25 years were recruited from the Johns Hopkins community. Of these, seventeen (mean age 20.4 years, 13 females, 17 right handed) participated in the written version of the experiment, and seventeen (mean age 20.2 years, 11 females, 14 right handed) participated in the spoken version of the experiment. Each participant received $60 upon completion of the final experimental session.

Stimuli

Participants were trained on a total of 24 novel items. Figure 12 shows an example stimulus; the complete set is shown in Appendix B.

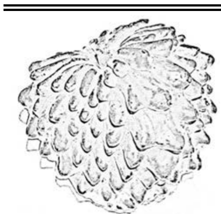
Picture	Written experiment spelling	Written experiment pronunciation	Spoken experiment spelling	Spoken experiment pronunciation	Features
	chys	JIs	crube	krub	Green Tastes sweet Used for food Consumed cooked

Figure 12. Example stimulus used in Study 1.

For each participant, there were six blocks of four items: two blocks of semantically related items, two blocks of segmentally related items, and two blocks of unrelated items. Across participants, each item appeared in only one type of block.

Pictures. For each item, a line drawing of a very unusual object with which participants were unlikely to be familiar was paired with a pseudoword name. Picture stimuli were taken from the Ancient Farming Equipment stimuli, the NOUN database (Horst, 2009), clip art directories, and images freely available online, which were modified using Adobe Photoshop CS5 to look like black and white line drawings and be approximately 4-inches by 4-inches in size. The same images were used in the written and spoken versions of the experiment.

Semantic content. Theories of semantic representation suggest that living things and artefacts have different organizations of shared and distinctive features, with living things having more shared features and artefacts having more distinctive features (e.g., Tyler & Moss, 2001). Furthermore, different types of features are differentially important for different categories (e.g., for review see Martin & Caramazza, 2003), regardless of their shared or distinctive status. Some theories of semantic representation suggest that sensory properties regarding the physical characteristics of an object (e.g., sweet taste) are more important for living things while functional properties regarding the use of an object (e.g., used for food) are more important for artefacts (e.g., Shallice & Warrington, 1984). The items used in this study included both living things and man-made artefacts. For each item, four semantic features were provided, consisting of two sensory features and two functional features. Features did not refer to visual properties of the object that were present in the black and white image so that the task required attention to the provided features.

Semantic overlap. In the semantically related blocks, feature overlap was distributed unpredictably such that each item had two features that were shared with at least one other item in the block and two features that were unique to the item (i.e., there were two shared and two distinctive features for each item). The same feature was never shared by all four items in a block. This distributed overlap was to ensure that participants had to learn which items had each feature: they could not merely assume that since an item was in a certain block it had a certain feature or infer that an item had a certain shared feature based on its other features (e.g. not all items that had a sweet taste were used for food). Within each segmentally related block and unrelated block, semantic features were not shared, since sharing semantic features would make the blocks semantically related. Semantic features were not repeated across segmentally related and unrelated blocks in order to minimize the semantic relatedness of these items. The same semantic features were assigned to the same images in the written and spoken versions of the experiment. These features were presented in written form for both the spoken and written experiments, including during the feature probe task.

Segmental content. Monosyllabic pseudoword names for the items were orthotactically and phonotactically plausible in English. Different pseudoword stimuli were used in the written and spoken versions of the experiment. All six blocks within a modality were matched on length in letters and phonemes of the pseudowords, as well as on measures of phonological and orthographic neighborhood density, including number and summed frequency of body, onset, and total orthographic neighbors and number and summed frequency of phonological neighbors; number and summed frequency of body friends and enemies; and position nonspecific and position specific type and token bigram and trigram frequency, using measures from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002) (these values are reported in Appendix C). Each pseudoword had four segments: four letters in the written version, and four phonemes in the spoken version.

Segmental overlap. In the segmentally related blocks, segment overlap was distributed unpredictably such that each item had two segments that were shared with at least one other item in the block and two segments that were unique to the item (i.e., there were two shared and two distinctive segments for each item). This context was chosen because, as described in Chapter 1 (Background Experiments 3 and 4), unpredictably distributed segmental overlap leads to interference during blocked cyclic naming of known items. Similar to the semantic manipulation, using this type of overlap meant that participants had to learn which items had each segment: they could not merely assume that since an item was in a certain block it had a certain segment. Shared segments did appear in the same position across items; distinctive segments did not appear elsewhere in the block in any position. Within each semantically related and unrelated block, each segment appeared only once in any position, meaning the segments in these words were distinctive and the blocks were not segmentally related. However, because there is a limited set of segments in English, segments necessarily repeated across blocks. This repetition across blocks was controlled so that each segment that appeared in a segmentally related block also appeared three times across the other blocks. Application of these constraints resulted in the same segment serving as a shared segment across both segmentally related blocks: this occurred for one shared segment in the written version and for two shared segments in the spoken version.

Encouraging lexical processing. In both the spoken and written versions of the experiment, the names of the items were presented both visually and auditorily during the familiarization phase of the experiment. This was to make it clear that there is not a consistent relationship between orthography and phonology. Participants were to rely on the orthography of the words, not the phonology, in the written version of the experiment, but vice versa in the spoken version. Although blocks were not formally matched on phoneme-to-grapheme or grapheme-to-phoneme regularity, some names shared letters but not sounds in the written condition (e.g., lisk and rish both share the letter s, but it is pronounced differently), and some names shared sounds but not letters in the spoken condition (e.g., crube and scoon share the same vowel but it is spelled differently). This manipulation was intended to reduce participant reliance on the sublexical route of production and encourage use of the lexical route. The manipulation should minimize concerns that participants in the written version of the experiment were learning a phonological form and then directly converting that to orthography via phoneme-to-grapheme conversion processes (or that participants in the spoken version of the experiment were learning an orthographic form and then directly converting that to phonology via grapheme-to-phoneme conversion processes). Encouraging lexical processing by demonstrating the inconsistency of spelling and pronunciation of the words should lead to learning of orthographic lexical representations in the written version of the experiment and phonological lexical representations in the spoken version. Note that in creating stimuli that respected the constraints of unpredictably distributed overlap and some inconsistencies between spelling and pronunciation, it was not possible to use the same pseudowords for the written and spoken versions of the experiment. In the written version, names had four letters that corresponded to three to four phonemes. In the spoken version, names had four phonemes that corresponded to five letters.

Recordings. All of the pseudowords were recorded by the same male speaker, who was a native speaker of American English. Recordings were normalized in length to 900 msec by adding silence to the beginning of each file. The same speaker also recorded sounds for the sound probe task, which consisted of each phoneme present in the spoken version of the experiment. Vowels were produced in isolation. Consonants were produced followed by /a/, a vowel that did not appear elsewhere in the experiment, since consonant sounds cannot be produced in isolation. These sound recordings were normalized in length to 625 msec. Letters in the letter probe task were presented visually.

Procedure

The experiment was run using E-Prime 2 Professional (Psychology Software Tools, Pittsburgh, PA) on a Dell Latitude E6500 laptop with a 13-inch by 8-inch screen. Participants attended four separate training sessions within 10 days, as well as a fifth follow up session approximately 2 weeks later (12-17 days later). The general procedures for familiarization and training were based on previous reports using the Ancient Farming Equipment paradigm (Grönholm-Nyman et al., 2010; Grönholm et al., 2005; Tuomiranta et al., 2011, 2012), although modifications were introduced to apply it to both the written and spoken modalities, to compare different blocking contexts, and to encourage retrieval attempts during training.

Familiarization phase. At the beginning of the first session, participants were familiarized with the twenty-four items to ensure that the items were initially unknown to the participants and thus that differences after training were due to that training, not to differences in pre-existing knowledge. First, all pictures were shown to the participants without any additional information. If a participant indicated that they knew the name of the picture, which was possible since they depicted real, albeit uncommon, objects, that picture was switched with a substitute from the same semantic category.

Next, names for the items were introduced to the participants. Here, items were presented in the same order for all participants, which preserved block structure (i.e., the four items from each block appeared in immediate succession). Participants in both the written and spoken versions of the experiment saw each 4-inch by 4-inch picture in the left top quarter of the screen. Next to the picture on the screen, a list of its four semantic features was printed in Arial size 18 font. Beneath these, the written spelling of the assigned name was printed in Arial size 24 font. Participants also heard an audio recording of the item name presented via headphones. After listening to the name and reading the information on the screen to themselves, they were instructed to press a button to advance to the next trial. In the written version of the experiment, this button was on the Wacom Bamboo graphics tablet that they used to write responses during other parts of the experiment. In the spoken version of the experiment, this button was on E-Prime's SR Box. Participants in the spoken version of the experiment were asked to repeat the name of the picture aloud to ensure that they heard it correctly. If participants did not press a button, the experiment automatically advanced to the next trial after 15 seconds. As described above, seeing the written form of the names and hearing the spoken form of the names was intended to encourage lexical processing since the phonology and orthography of the words were not completely consistent. The point of the familiarization phase was to provide information about the experimental stimuli so participants could begin the training task.

Instructions. Following the item familiarization phase, participants were given instructions about how to respond to naming trials throughout the experiment.

Written experiment instructions. In the written version of the experiment, participants were instructed on the use of the tablet. At the beginning of each trial, they were to place the pen on a marked starting point on a non-responsive edge of the tablet approximately 1 inch below the writing surface. As soon as they knew the name of the picture, they were to begin writing in a 2.5-inch by 1-inch rectangle centered at the bottom of the responsive surface. Response time (RT) was recorded when the writing surface was touched, and the participant's pen strokes appeared on the monitor. Participants were instructed to write as legibly as possible, which was made difficult by the fact that pen movements were recorded on the screen even when the pen was lifted between letters. Suggestions for improving legibility were given, including trying to pick up the pen as little as possible, crossing t's as soon as they were written, not dotting i's, and looking at the writing on the screen instead of at the tablet. It was emphasized that they should *begin* writing as soon as possible, but that they did not have to carry out the writing of the whole name quickly. However, they should not begin until they knew what they wanted to write since the picture disappeared when the writing surface was touched. Participants were instructed to write what they knew about the names, even if they were not sure of all the letters. They were to write the letters they knew and draw blanks for unknown letters (e.g., for a name that started with c and ends in s, they could write c_ _ s). They were to draw a line through the box if they could not remember any of the letters. After writing a response, participants were to return the pen to the marked starting point and press a button with their non-dominant hand to advance to the next trial. When this button was pressed, a screen shot of the completed response was saved, which was used to score accuracy.

Spoken experiment instructions. In the spoken version of the experiment, participants were instructed on the use of the microphone and SR-BOX. As soon as they knew the name of the picture, they were to say the name loudly and clearly into the microphone. They were instructed to avoid adding any extra words or sounds (such as "um", "it's a…", lip smacking, or throat clearing) because such sounds triggered the microphone. Response time (RT) was recorded when the microphone was triggered. As in the written version, they were to provide any information that they had about the names, even if that information was incomplete (e.g., they could say that something started with "ba" or rhymed with "cat"). They were told to say, "I don't know" if they did not remember anything about the sounds of the name. After saying their response, participants were to press a button on the SR-BOX to advance to the next trial. Responses were recorded with a separate audio recorder. These recordings could be compared to the accuracy scoring conducted on-line during the experiment by the experimenter.

Training phase.

Trial structure. Each trial consisted of five parts. The structure of one complete written trial is depicted in Figure 13.

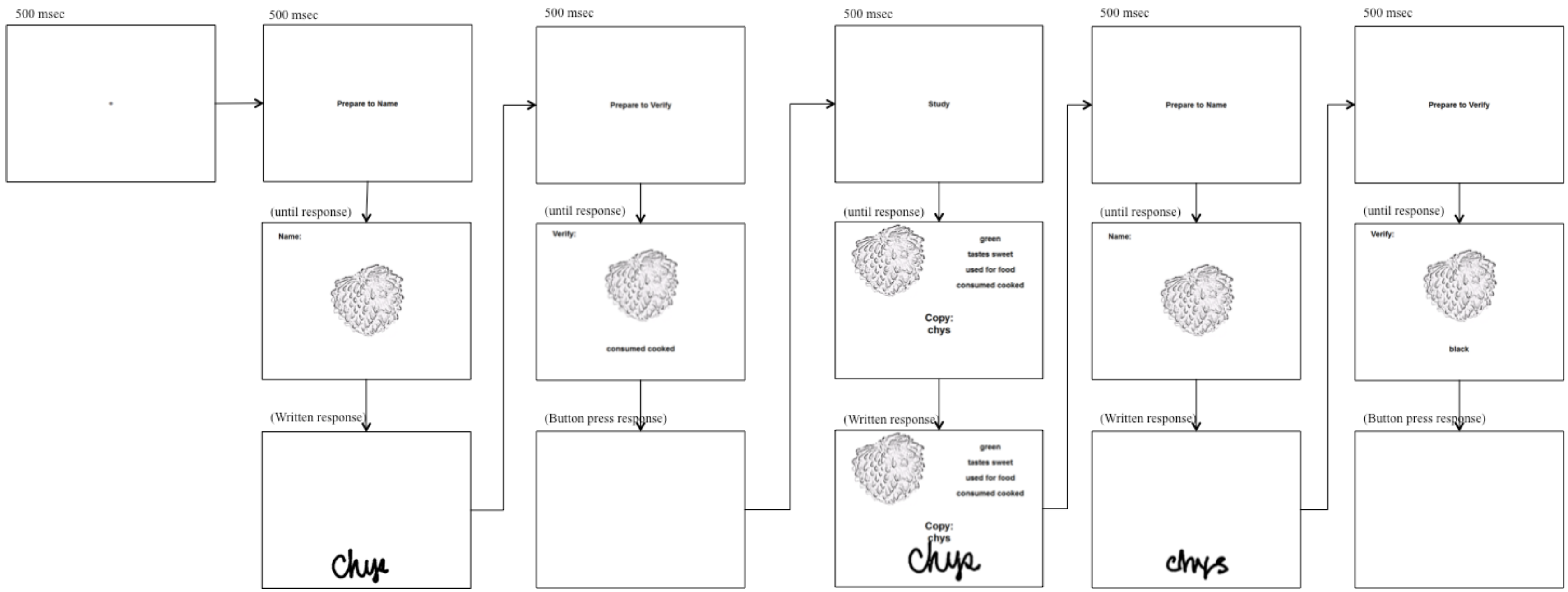


Figure 13. Structure of written trial from the training task in Study 1.

Trials in the spoken experiment differed in that responses were said aloud instead of written and in that the study phase, the name of the item was auditorily presented instead of printed on the screen.

First, following a 500-msec presentation of a fixation cross, “Prepare to Name” appeared on the screen in Arial size 18 font for 500 msec. During the naming part of the trial that followed, participants saw a 4-inch by 4-inch picture in the center of the screen and attempted to name it as soon as possible. The instruction “Name:” appeared at the top left of the screen in Arial size 18 font. The picture remained on the screen until the response was initiated; there was no time limit for beginning a response although participants were instructed to respond as soon as possible. At that point, the picture disappeared from the screen, and participants pressed a marked button to continue to the next trial when they had completed their responses.

Second, “Prepare to Verify” appeared on the screen in Arial size 18 font for 500 msec. During the verify part of the trial that followed, participants saw the same 4-inch by 4-inch picture in the center of the screen, accompanied by a written semantic feature printed underneath the picture in Arial size 18 font, as well as the instruction “Verify:” at the top left of the screen in Arial size 18 font. Participants were to press a button marked YES if the item had that semantic feature and a button marked NO button if it did not. In the written version, this button pressing was performed with the non-dominant hand as participants continued to hold the pen with the dominant hand. In the spoken version, participants were not constrained to use of the non-dominant hand. Again, there was no time limit for responses in this phase of the experiment, but participants were instructed to respond as soon as possible.

Third, “Study” appeared on the screen in Arial size 18 font for 500 msec. During the study part of the trial that followed, participants again saw the same 4-inch by 4-inch picture, now in the top left quarter of the screen. Next to it, all four features of the item were printed in Arial size 18 font. In the written version of the experiment, the name of the picture along with the instruction “Copy:” was printed in Arial size 24 font. Participants were to review the information and copy the name of the picture. Nothing on the screen disappeared when the copy response was written; the response appeared on the screen under the word “Copy”. In the spoken version of the experiment, the instruction “Repeat:” was printed in Arial size 24 font, and a recording of the name was played once over the headphones. Participants repeated the name aloud. If they did not pronounce the name correctly (e.g., pronouncing “bloof” as “blooth”), the experimenter corrected them at this point. Note that, unlike in the familiarization phase, participants in the written version of the experiment saw the name but did not hear it, while participants in the spoken version of the experiment heard the name but did not see it. In the study part of the trial, all information remained on the screen while the participant responded until they pressed a button to continue to the next part of the trial. Participants were told that they did not need to respond as quickly as possible but could spend as much time as they wanted on this part of the task.

After the study portion of the trial, participants again attempted to name the same picture and verify a different semantic feature as quickly as possible, following the same procedures as in the first two parts of the trial. That is, after the first three parts of the trial, parts one and two were repeated. Overall, the five parts of the trial made up a test-study-test procedure, with participants first being tested on the name and semantic features of a picture, then studying that information, and finally being tested on it again.

There was a 500 msec blank inter-trial interval between trials.

Practice phase. During the first session, participants practiced the task by completing five trials with known words (bird, hose, belt, rose, and sled). This practice allowed them to become comfortable with the many parts of the task, and to practice using the response apparatus appropriately.

Training task. Following the practice phase, participants began the training task, which consisted of 72 training trials of the stimuli they were learning in the experiment. Each trial followed the structure defined above. Trials were presented in blocks. One block of each type (segmental, semantic, and unrelated) was presented before repetition of any block type. Participants completed training on four days: the pseudorandom order of blocks was never repeated within participant. Within each block, all four items were presented over three cycles in random order for a total of 12 training trials per block. For the verify portions of the trials, all incorrect features consisted of correct features for other items from the same block. There were two opportunities to verify features in each trial, and the same feature was never repeated within the same trial. Over all sessions for the 17 participants, features to verify were correct for 50% of trials (range of 46%-54% in a single session) and incorrect on 50% of trials (range of 46%-54% in a single session). Each feature was repeated as a correct feature to verify 1-11 times and as an incorrect feature 1-10 times over the course of the four training sessions for each participant (each individual shared feature appeared more times than each distinctive feature since shared features were correct for more than one item).

Evaluation phase. After training, the evaluation phase consisted of the recall task and the distinctiveness probe tasks (semantic feature probe task and segment probe task), which assessed learning of the items. Feedback was not given on any of these tasks.

Recall task. In the recall task, participants saw all twenty-four pictures individually in a random order and were asked to produce the name for each as quickly and accurately as possible. Following a 500 msec presentation of a fixation cross, a 4-inch by 4-inch picture was presented in the center of the screen. This disappeared as soon as participants initiated their response. There was no time limit for responding. As in the naming part of the training trials, they pressed a marked button to continue to the next trial. In the written version of the experiment, there was a cumulative inter-trial interval of 3000 msec, meaning the next trial began a minimum of 3000 msec after the previous trial. In the spoken version of the experiment, this was shortened to 1000 msec because response duration is substantially shorter when speaking as opposed to writing. As in training, responses were recorded for use in scoring accuracy via screen shots in the written version of the experiment or via audio recording in the spoken version of the experiment.

Semantic feature probe task. The semantic feature probe task began with a 500 msec fixation cross. Next, a 4-inch by 4-inch picture appeared in the center of the screen, below which was printed a semantic feature in Arial size 18 font. Participants were to decide if the item in the picture had the given semantic feature or not by pressing the same YES and NO buttons that were used in the verify parts of the training trials, and response time was recorded. Participants had 2000 msec to respond before the experiment automatically advanced to a 500 msec inter-trial interval and then to the next trial.

In all probe tasks, participants were instructed to use their index fingers to press the buttons. In the written version of the experiment, the tablet was physically rotated so that the buttons appeared at the bottom of the tablet instead of on the side of the non-dominant hand. Half the trials had YES responses, while half had NO responses. There were 192 trials: there was a YES trial for each of the four features for each item, as well as four NO trials for each item that paired each picture with an incorrect feature that was true for another item trained in the same block.

Segment probe task. Finally, participants completed a segment probe task. While the training task included components that were like the recall task and semantic feature probe task, only the evaluation phase included a segment probe component.

In the written version of the experiment, the segment probe task was a letter probe task. The set-up of the letter probe task was the same as the semantic probe task, except that a single upper-case letter was presented in Arial size 18 font instead of a feature. The task was to press the YES button if the name of the item contained the letter and the NO button if it did not. Again, participants had 2000 msec to respond. There were 192 trials, half of which had YES responses that correctly matched each picture with each of its four letters and half of which had NO responses that matched each picture with four other letters that appeared in other items trained in the same block.

In the spoken version of the experiment, the segment probe task was a sound probe task. Sounds were played over headphones in place of the letters that were displayed visually on the screen in the letter probe task. Participants were to decide if the sound played appeared in the name of the picture displayed on the screen by pressing buttons as in the semantic feature probe and letter probe tasks. This task was made more difficult by the fact that consonants cannot be produced in isolation. In these recordings, the phoneme /a/ was used with all consonants. Participants were instructed to focus on the consonant, ignoring the /a/ that did not appear in any of the items on which they were trained. To compensate for the additional time that it took to play the sounds, the time limit was increased to 2650 msec for the sound probe task.

Session structure. The first experimental session consisted of a familiarization task, practice with the training task, a training task, a recall task, a semantic feature probe task, and a segment probe task. The second, third, and fourth sessions followed a similar order, with the exception that the familiarization and practice tasks were replaced by another administration of the recall assessment in order to see what was retained from the previous session (i.e., the order was recall task, training task, another administration of the recall task, semantic feature probe task, and finally segment probe task). While tasks were repeated in the same order across sessions, the order of items in the tasks differed for each administration. After the four training sessions were completed, there was a retention interval of approximately two weeks. Participants then returned for a fifth session in which they completed the recall task, semantic feature probe task, and segment probe task. The results of this follow-up session were used to assess long-term learning of the items.

Results and Discussion

The purpose of this study was to test the e-ILM predictions presented in the previous chapter. In this section, the results that address each of the predictions listed in Table 4 are presented in turn.

All data presented in this chapter were analyzed using multilevel mixed models with random effects in R version 3.2.4 with the lme4 package (version 1.1-12). The presented *p*-values were calculated using the approximation of the normal distribution, which was applicable to both the generalized linear mixed effects models used for accuracy data and the linear mixed effects models used for response time data. Accuracy data were analyzed using logistic regression as they are binary, while log-transformed response time data were analyzed using linear regression. Since the predictions all consider the impact of a specific type of blocking (semantic or segmental) relative to an unrelated context, separate models were constructed to compare items trained in semantic vs. unrelated contexts and items trained in segmental vs. unrelated context. Items trained in semantic and segmental blocks are not directly compared to one another since this comparison is not relevant for testing the predictions derived from the e-ILM.

Training

The first e-ILM predictions concern effects during training in blocked contexts. According to the predictions, both semantic blocking and segmental blocking should lead to interference during acquisition relative to training in unrelated contexts. If the predictions are correct, this increased interference should be observable as reduced accuracy and/or increased response times during training for items trained in each blocked context relative to items trained in unrelated contexts. In order to evaluate the predictions, data collected during the training task were analyzed.

Overall performance on training task. Written and spoken data were analyzed separately. Training analyses considered only the first naming attempt made in each trial. Participants performed at ceiling on the other portions of the trials, with 99.8% whole response accuracy on written copy, 97.4% accuracy on spoken repetition, 99.6% accuracy on written second naming attempt, and 98.8% accuracy on spoken second naming attempt, indicating that participants were attending to the task and capable of performing it.

Overall, across all training sessions, participants correctly produced the whole response on 77.5% of first naming attempts in the written experiment and on 71.3% of first naming attempts in the spoken experiment. The majority of errors were omissions in which no segments were produced (66.8% of errors in the written experiment and 71.7% of errors in the spoken experiment). There were a few within-block substitutions whereby another name from the same block was produced instead of the target (5.1% of errors in the written experiment and 2.6% of errors in the spoken experiment) and across-block substitutions whereby another name from the experiment that was not trained in the same block was produced instead of the target (3.7% of errors in the written experiment and 3.1% of errors in the spoken experiment). Remaining errors consisted of additions, deletions, and substitutions of segments (24.3% of errors in the written experiment and 22.6% of errors in the spoken experiment). Accuracy analyses examined whole response accuracy, not segment accuracy, a reasonable choice given that only 5.4% of trials in the written experiment and 6.4% of trials in the spoken experiment were not correct responses, whole omissions, or whole response substitutions.

Training analysis structure: Accuracy models. Logistic regression models examined whole response accuracy for first naming attempts during the training task. Written and spoken data were analyzed in different models. Within each modality, separate models were constructed to directly compare each individual blocking context to the unrelated context. The semantic models excluded data from the segmental context, while the segmental models excluded data from the semantic context. This means there were four total accuracy models: written semantic, written segmental, spoken semantic, and spoken segmental.

These training accuracy models had the same basic structure, including the following fixed effects: block type (semantic, segmental, or unrelated context), training attempt within session (1-3), training session (1-4), two- and three-way interactions between those, and the control variables of days since last training session and number of training trials since the target was last trained. Continuous variables were centered and scaled. In the semantic model, block type was coded as semantic=1, unrelated=-1. Likewise, in the segmental model, block type was coded as segmental=1, unrelated=-1. A full random structure was implemented in each model, with random intercepts for subjects and items, a full random slope structure matching the fixed effect structure (block type, training attempt within session, training session, two- and three-way interactions between those, days since last training session, number of training trials since the target was last trained) over subjects, and the same random slope structure over items with the exception that block type and its interactions was excluded since each item was trained in only one context.

Explanation of included variables. In order to focus the analysis on effects during training, both training attempt within session and session were included as variables. An alternative choice, using training attempt number across sessions, would have added complications due to conflating recall and training. At the beginning of each session, there would likely be a dip in accuracy compared to the previous attempt because the item must be recalled from the last session. Changes in recall (if, for instance, words trained in a semantic block are recalled better than words trained in an unrelated block) might impact the apparent effects of training, masking effects that are in fact due to training differences or revealing effects that are actually due to recall differences. Using training attempt within session allows focus on what happens during training. A positive main effect of training attempt would show that participants improve as they practice naming the same item multiple times in the same session.

Adding session as a variable allows investigation of whether this effect changes as more sessions are completed as would be seen via the two-way interaction of training attempt within session and session, as well as demonstrating learning over the course of the experiment as would be seen via a positive main effect of session.

The main question of interest is whether training in a blocked context increases difficulty. In order to address this question, the two-way interaction between block type and training attempt within session was used. If blocking increases difficulty as predicted, a negative interaction is expected whereby blocked items improve less within session than unblocked items. This two-way interaction of block type and training attempt within session was more appropriate for addressing the question than the two-way interaction of block type and session because the interaction of block type and session is influenced by possible differences in retention for blocked vs. unrelated items across sessions, not just by differences in training effects within sessions. For example, blocking might make training itself more difficult but make retaining the information until the next session easier, and both of these effects would impact the interaction of block type and session. The interaction of block type and training attempt, however, specifically examines effects of blocking within each session, allowing investigation of the effects of blocking during training but not retention across sessions.

The three-way interaction of block type, training attempt within session, and session examines whether this effect changes over time as participants learn more. For example, items in unrelated blocks may experience little interference at the beginning of training, while those in related blocks may experience some interference; so items in unrelated blocks may improve more quickly than items in related blocks especially during early training sessions.

The control variables of days since last training session and number of training trials since the target was last trained evaluate whether accuracy and speed change as a result of the spacing of training, where spacing is measured by time or amount of intervening information.

Training analysis structure: Response time models. Analysis of training data was not limited to accuracy. Four models (written semantic, written segmental, spoken semantic, and spoken segmental) were also constructed to evaluate effects on response time (RT). Response times for inaccurate responses were excluded, as were outliers more extreme than 2.5 standard deviations from each participant's mean RT regardless of accuracy. In the analysis of the spoken experiment, response times for trials in which the microphone was incorrectly triggered, either prematurely or belatedly, were also excluded. Log-transformed response times were analyzed using multilevel mixed models with random effects. In this case, linear regression was appropriate for the data as opposed to the logistic regression used for the accuracy data reported above. The model architecture was otherwise the same as in the accuracy analysis. Difficulty increases correspond to RT increases as opposed to accuracy decreases; therefore, the directions of the predictions discussed for accuracy are reversed when considering response time.

Written experiment training results. Figure 14 shows the results of the written training task. The results tables of analyses presented in this chapter are presented in full in Appendix D. Tables D1 and D2 show the results of the semantic and segmental models of training accuracy and response time data from the written experiment. Significant ($p<.05$) and marginally significant ($.05<p<.1$) effects are discussed in the text below.

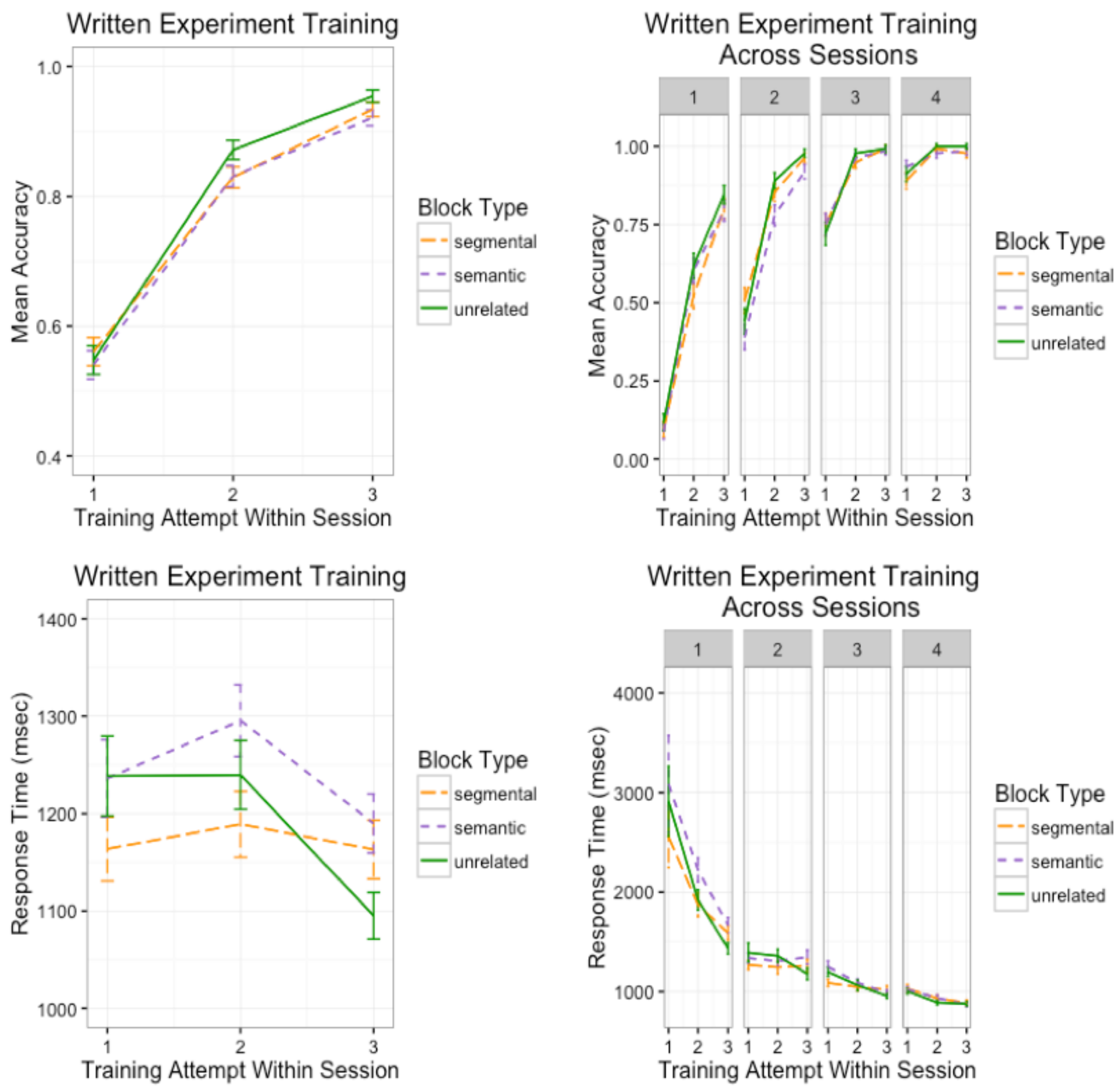


Figure 14. Results of the training task from the written experiment.

The top left panel shows accuracy across the three training attempts within each session, collapsed across all four training session for the three training contexts.

The top right panel shows accuracy across the three training attempts within each session for each of the four training sessions for the three training contexts.

The bottom left panel shows response time across the three training attempts within each session, collapsed across all four training session for the three training contexts.

The bottom right panel shows response time across the three training attempts within each session for each of the four training sessions for the three training contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

The critical effect to examine how context impacted training difficulty was the interaction of block type and training attempt. For the semantic model of the written experiment, there was a negative interaction between block type and training attempt within session for accuracy (z=-2.21, p=.027) such that participants had smaller increases in accuracy across attempts for items trained in semantic contexts than for items trained in an unrelated context. The interaction for response time did not reach significance (t=1.22, p=.221), although its direction was consistent with increased difficulty in terms of smaller increases in speed (smaller decreases in response times) for items trained in semantic vs. unrelated contexts. For the segmental model of the written experiment, although the interaction for accuracy did not reach significance (z=-1.54, p=.125), it was again in the direction of increased difficulty in terms of smaller increases in accuracy for items trained in segmental vs. unrelated contexts. In this model, there was a positive interaction between segmental vs. unrelated block type and training attempt within session for response time (t=3.20, p=.001) such that participants had smaller increases in speed (smaller decreases in response time) across attempts for items trained in segmental contexts than for items trained in an unrelated context. Overall, these results for both semantic and segmental blocking are in accordance with the e-ILM predictions: blocking increased difficulty during training, with items trained in a blocked context improving more slowly across training attempts within session than items trained in an unrelated context.

A number of other significant effects were also observed in the analysis of written training data. First, participants did learn the items throughout the experiment. There were consistent main effects of session such that participants increased in accuracy

(z=7.43, p<.001 for the semantic model; z=8.00, p<.001 for the segmental model) and decreased in response time (t=-16.24, p<.001 for the semantic model; t=-15.94, p<0.001 for the segmental model) as they completed more training sessions, indicating learning across the course of the experiment. Furthermore, there were also consistent main effects of training attempt within session such that participants increased in accuracy (z=6.11, p<.001 for the semantic model; z=5.10, p<.001 for the segmental model) and decreased in response time (t=-6.53, p<.001 for the semantic model; t=-5.91, p<.001 for the segmental model) as they practiced naming the same item multiple times within a session, again demonstrating learning. Note that there was a significant interaction between the two effects of session and training attempt within session for response time in the semantic model (t=2.63, p=.008), indicating that the decrease in response time across attempts within a session was reduced as training proceeded, which may have been due to better performance in later sessions leaving less room for improvement. However, the interaction was not significant when examining accuracy in the semantic model, or when examining accuracy or response time in the segmental model.

Second, there were several other indications of increased difficulty due to training in semantic and segmental blocks as compared to unrelated contexts. In terms of the main effect of block type, there was a significant accuracy disadvantage for the semantic as opposed to unrelated context (z=-2.08, p=.038) in the semantic model, and a marginally significant accuracy disadvantage for the segmental as opposed to unrelated context (z=-1.81, p=.070) in the segmental model. This suggests that training in a blocked context led to difficulty throughout training. There were not significant effects of block type on response time. Further support for the negative impact of block type on training comes from the interactions of block type and session. There was a significant negative interaction for semantic vs. unrelated block type and session for accuracy in the semantic model (z=-1.96, p=.050). However, there was not a significant interaction for segmental vs. unrelated block type and session for accuracy in the segmental model. There were no significant interactions of block type and session on response time. It is possible to further characterize the effects by examining the three-way interactions of block type, training attempt within session, and session. There was a marginally significant negative three-way interaction for accuracy in the semantic model (z=-1.74, p=.081), though not for the segmental model. The three-way interactions were negative and significant when examining response time in the semantic model (t=-1.97, p=.049) and in the segmental model (t=-2.57, p=.010). These interactions suggest that participants showed more improvement within a session for blocked contexts than unblocked contexts as they completed more training sessions. This may be because participants improve more quickly for the unrelated contexts in early sessions and then have less room for improvement in later sessions, while they improve more slowly for the blocked contexts and thus continue to improve in later sessions.

Finally, the variables of days since the last session and training trials since last trained were included to examine the effects of spacing on training. There were no significant effects of days since the last session. There was a negative effect of trials since last trained on response time in the semantic model (t=-2.12, p=.034), indicating that spacing during training led to an increase in speed of production. This suggests that spacing in terms of interleaving trials may be beneficial for performance during training.

Spoken experiment training results. Figure 15 shows the results of the spoken training task. Tables D3 and D4 (in Appendix D) show the results of the semantic and segmental models of training accuracy and response time data from the spoken experiment.

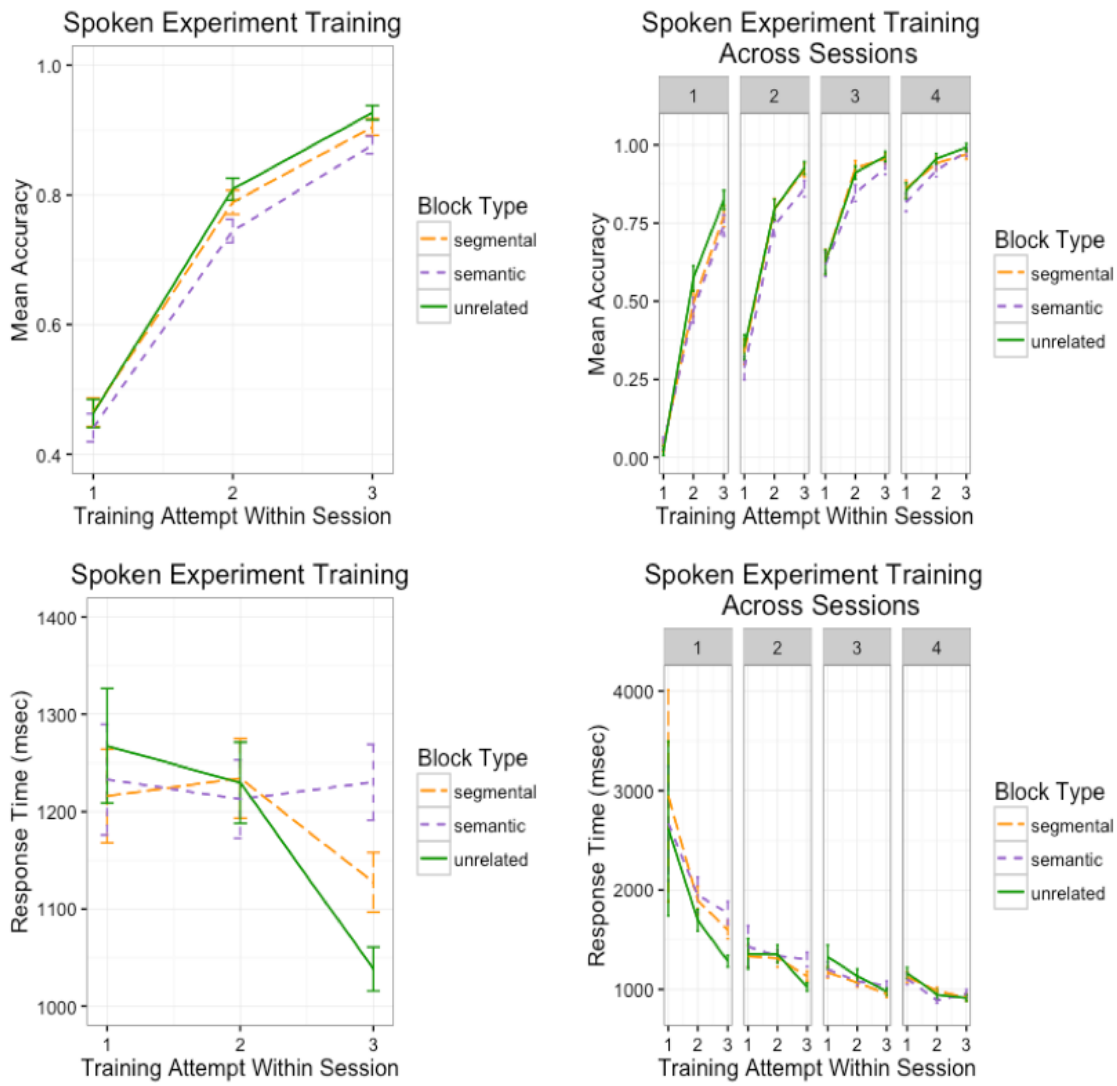


Figure 15. Results of the training task from the spoken experiment.

The top left panel shows accuracy across the three training attempts within each session, collapsed across all four training session for the three training contexts.

The top right panel shows accuracy across the three training attempts within each session for each of the four training sessions for the three training contexts.

The bottom left panel shows response time across the three training attempts within each session, collapsed across all four training session for the three training contexts.

The bottom right panel shows response time across the three training attempts within each session for each of the four training sessions for the three training contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

In the spoken experiment, as in the written experiment, the critical test of the hypothesis that blocking increases training difficulty is the interaction between block type and training attempt within session. For both the semantic and segmental models, this two-way interaction was significant for response time (t=2.34, p=.019 for the semantic model; t=2.00, p=.046 for the segmental model), indicating that participants showed less improvement in speed (smaller decreases in response time) over practice trials of the same item within training sessions for items trained in semantic vs. unrelated and segmental vs. unrelated context. The interaction effects for accuracy did not attain significance (z=-0.85, p=.393 for the semantic model; z=-0.92, p=.355 for the segmental model). The significant positive interactions for response time are in line with the e-ILM predictions: training in blocks leads to interference during acquisition relative to training in unrelated contexts, both when those blocks are semantically similar and when they are segmentally similar.

Again, other significant effects were also observed in this analysis. As expected, participants did learn throughout the experiment. Evidence for this is derived from the main effects of session on accuracy (z=10.78, p<.001 for the semantic model; z=10.99, p<.001 for the segmental model) and response time (t=-7.43, p<.001 for the semantic model; t=-6.51, p<0.001 for the segmental model), indicating improvement as more training is completed. Participants also demonstrated learning within sessions, as shown by main effects of training attempt within session on accuracy (z=7.16, p<.001 for the semantic model; z=7.79, p<.001 for the segmental model) and response time (t=-3.36, p<.001 for the semantic model; t=-5.10, p<0.001 for the segmental model). There was a significant interaction between the two effects for accuracy in both models (z=-2.39,

p=.017 for the semantic model; z=-1.99, p=.046 for the segmental model), which showed that that the increase in accuracy across attempts within a session was reduced as training proceeded, which may have been due to better performance from the beginning of later sessions which allows less room for improvement. This two-way interaction of session and training attempt within session was not significant for response time.

In contrast to the written experiment, there were few additional indications of increased difficulty due to blocking from other effects in the spoken experiment models. There were no main effects of block type on either accuracy or response time, indicating that there was not a consistent advantage or disadvantage for items trained in any of the three blocking contexts. In the semantic model, there was a significant negative interaction between block type and session for response time (t=-2.31, p=.021), although the effect was not significant for accuracy in the semantic model or for either response time or accuracy in the segmental model. This suggests that items trained in a blocked context, especially a semantic one, improved more in speed (showed greater decreases in response time) across sessions than did unrelated items. This is consistent with the idea that unrelated items improved more quickly (i.e., in earlier sessions) and had less room for improvement in later sessions than did blocked items which took more sessions to learn. The three-way interaction between block type, training attempt within session, and session was not significant for accuracy or response time in any model, indicating that the effect of block type on improvement within sessions did not change across sessions.

Additionally, the effects of spacing on training were examined via the variables of days since the last session and training trials since last trained. There was a marginally significant negative effect of days since the last session on accuracy in the semantic model (z=-1.93, p=.054), and a significant positive effect of days since the last session on response time in the segmental model (t=2.43, p=.015). Together, these results show that increasing the time between training sessions was detrimental to performance, indicating that this type of spacing increased difficulty. There were no significant effects of training trials since the target was last trained, indicating that the number of intervening items did not clearly impact performance. Note that these results do differ from those found in the written experiment, in which there were no significant effects of days and a negative effect of trials since last trained for response time in the semantic model. The contrasting effects suggest that spacing effects require further investigation.

Written and spoken experiment training results. Overall, in the spoken model as in the written model, the blocked contexts had negative effects on training. While participants did learn over the course of the experiment, improvement was slower within sessions for both blocking contexts in both experiments. This is consistent with the prediction that training in both semantic and segmental contexts increases interference during acquisition. Although all models had numerical effects of increased interference in blocked as opposed to unrelated contexts, there were some differences in significant effects between the written and spoken experiments: the interaction between semantic vs. unrelated block type and training attempt within session appeared as a significant effect for accuracy in the written model and for response time in the spoken model. This difference, as well as other differences in the simple effects of block type and its other interactions, may be due to differences in the time courses of production: spoken production takes place over a shorter time course and in a more parallel fashion than written production. The main finding, that training increases difficulty during training as predicted by the e-ILM, is consistent across modalities.

Distinctiveness

The second set of e-ILM predictions concern distinctiveness changes as a result of training in blocked contexts. According to the predictions, semantic blocking during training increases the distinctiveness of representations by strengthening the connections between distinctive semantic features and lexical nodes while weakening the connections between shared semantic features and lexical nodes. On the other hand, segmental blocking during training is predicted to reduce the distinctiveness of representations by strengthening the connections between lexical nodes and shared segments while weakening the connections between lexical nodes and distinctive segments. In order to evaluate these predictions, the accuracy and speed of responses collected in the distinctiveness probe tasks were analyzed. If semantic blocking enhances distinctiveness as predicted, an advantage for distinctive relative to shared semantic features is expected in the semantic probe task. If segmental blocking reduces distinctiveness as predicted, an advantage for shared relative to distinctive segments is expected in the letter and sound probe tasks.

Overall performance on distinctiveness probe tasks. Participants were able to complete the distinctiveness probe tasks, although performance never reached ceiling. On the written semantic probe task, over all five administrations, participants correctly accepted features on 84.1% of trials (low of 77.0% on session 1 to high of 88.5% on session 4) and correctly rejected features on 83.9% of trials (low of 81.4% on session 1 to high of 86.7% on session 4). On the spoken semantic probe task, they correctly accepted features on 86.3% of trials (low of 79.4% on session 1 to high of 90.6% on session 4) and correctly rejected features on 87.6% of trials (low of 82.1% on session 1 to high of 93.2% on session 4). On the written segmental letter probe task, they correctly accepted segments on 72.4% of trials (low of 55.6% on session 1 to high of 84.7% on session 4) and correctly rejected segments on 85.9% of trials (low of 71.3% on session 1 to high of 92.5% on session 4). On the spoken segmental sound probe task, they correctly accepted segments on 69.4% of trials (low of 52.3% on session 1 to high of 78.6% on session 4) and correctly rejected segments on 84.4% of trials (low of 75.5% on session 1 to high of 90.5% on session 4).

Distinctiveness analysis structure: General. In the distinctiveness analyses addressing the predictions regarding whether training in blocks enhances or diminishes distinctiveness, only responses to correct features were considered (i.e., yes and no responses to probes in which the correct response was yes) because there were clear predictions about the effects of distinctiveness on these. If distinctiveness is enhanced, participants should be faster and more accurate to accept distinctive features or segments than to accept shared features or segments. If distinctiveness is reduced, on the other hand, participants should be faster and more accurate to accept shared features or segments than to accept distinctive features or segments. It is less clear if rejection of incorrect features should follow the same pattern, especially because some rejections could be made on the basis of general knowledge before any training even occurred (e.g., a type of tree is not going to swim).

Semantic probe and segment probe data were considered separately. For each probe task, two comparisons of shared and distinctive features or segments were used to evaluate the predictions. First, responses to shared features or segments were compared to responses to distinctive features or segments within the same blocking context. That is, for the semantic probe task, responses to shared features, which only occur for items trained in semantic blocks, were compared to responses to distinctive features for those same items trained in semantic blocks. In the letter probe and sound probe tasks, responses to shared segments, which only occur for items trained in segmental blocks, were compared to responses to distinctive segments for those same items trained in segmental blocks. These comparisons directly test the hypotheses that training in blocks reduces or enhances distinctiveness by differentially weighting shared and distinctive features or segments. Second, responses to shared features or segments were compared to responses to features or segments from the other blocking contexts. That is, for the semantic probe task, responses to shared features, which only occur for items trained in semantic blocks, were compared to responses to features for items trained in segmental and unrelated blocks, all of which are distinctive. In the letter probe and sound probe tasks, responses to shared segments, which only occur for items trained in segmental blocks, were compared to responses to segments for items trained in semantic and unrelated blocks, all of which are distinctive. This comparison provided a larger sample of responses to distinctive features or segments to compare to the shared features or segments.

As in the analysis of training data, written and spoken data were analyzed separately, as were accuracy and response time. Response times entered into analyses were log-transformed and excluded incorrect responses and outliers more extreme than

2.5 standard deviations from each participant's mean (outliers were calculated using means for all trials including those where the correct response was no).

Distinctiveness probe tasks were analyzed over two time scales. First, I examined how distinctiveness effects develop throughout training by analyzing the probe task data collected at the end of each of the four training sessions. Second, I examined whether distinctiveness effects persist to follow-up by analyzing the probe task data collected two weeks after training was completed, providing a measure of longer-term effects on distinctiveness.

Distinctiveness across sessions. Distinctiveness analyses first considered the results of the probe tasks across the four training sessions.

Model structure. All analyses of distinctiveness changes across the four training sessions were carried out using multilevel mixed models with random effects that relied on the same architecture. There were sixteen total models to investigate effects of distinctiveness across sessions, one for each combination of modality (written or spoken), task (semantic probe or segment probe), data type (accuracy or response time), and distinctiveness comparison (comparing shared and distinctive features/segments trained in the same or different blocks).

Each model included feature type (shared or distinctive), session (1-4), and the two-way interaction between them as well as days since the last session as fixed effects. Feature type was a categorical variable coded as shared=-1, distinctive=1. Training session was treated as a continuous variable that was centered and scaled. Due to failures of convergence, it was not possible to include a full random effects structure. Therefore, random slopes over items were not included. The resulting random effects structure included random intercepts for subjects and items as well as random slopes over subjects for feature type, training session, their interaction, and days since the last session.

Explanation of included variables. In all these models, the critical effects for testing changes in distinctiveness over time were the interaction between feature type and training session as well as a main effect of feature type. According to the hypotheses presented above, training in semantic blocks should enhance distinctiveness. Therefore, in the semantic probe tasks, participants should become faster and/or more accurate at verifying distinctive features relative to shared features as training proceeds, which would be reflected in the interaction between feature type and training session. Contrastingly, the hypotheses suggest that training in segmental blocks should reduce distinctiveness. In this segment probe tasks, participants should become faster and/or more accurate at verifying shared features relative to distinctive features as training proceeds, which would again be reflected in the interaction between feature type and training session. Main effects of feature type would indicate that there are differences in responses to shared vs. distinctive features that persist throughout the entire experiment. This is possible since participants have begun training on the semantic features and segments prior to the first administration of the probe tasks. Finding significant main effects of feature type would also support the hypotheses about distinctiveness: participants are expected to be faster and/or more accurate at verifying distinctive features relative to shared features in semantic probe tasks, but faster and/or more accurate at verifying shared features relative to distinctive features in segment probe tasks.

Main effects of training session were included to evaluate performance over the course of the experiment. Positive main effects of training session would indicate that participants get better at the task with more training, demonstrating learning.

Days since the last session was also included to evaluate whether there is an effect of the spacing of training sessions in time on speed and accuracy of responses.

Semantic probe results over sessions. The results of the analyses of the written and spoken semantic probe tasks over the four training sessions are presented below.

Written semantic probe results over sessions. Figure 16 shows the results of the written semantic probe task over sessions. Tables D5 and D6 (in Appendix D) report the results of the analyses of this task.

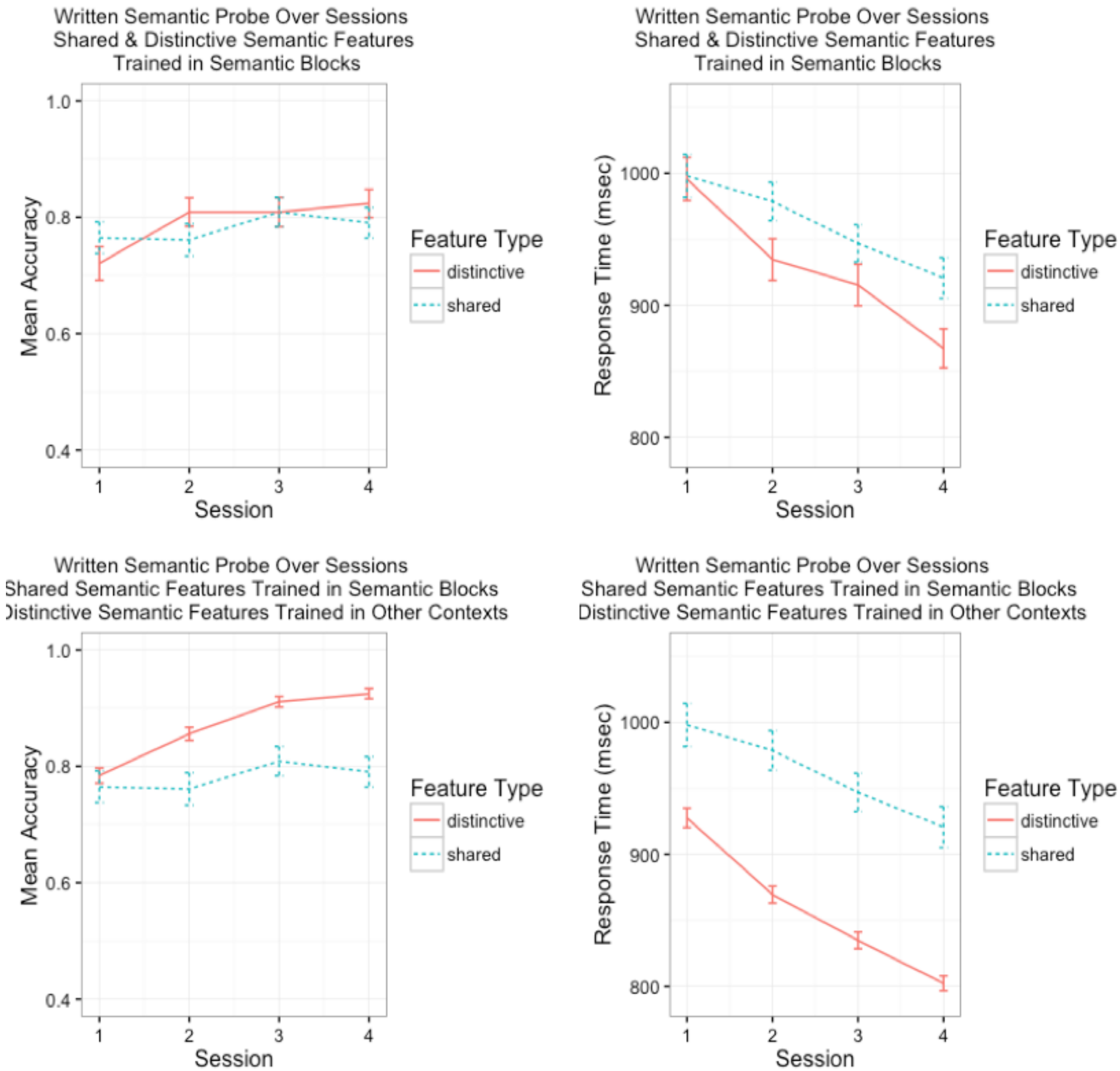


Figure 16. Results of the semantic probe task over sessions from the written experiment.

The top left panel shows accuracy for verification of semantic features across sessions, comparing shared and distinctive features trained in semantic blocks.

The top left right shows response time for verification of semantic features across sessions, comparing shared and distinctive features trained in semantic blocks.

The bottom left panel shows accuracy for verification of semantic features across sessions, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

The bottom right shows response time for verification of semantic features across sessions, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

According to the e-ILM predictions, training in semantic blocks should increase distinctiveness. To evaluate this prediction, the main effect of feature type and the two-way interaction of feature type and session were examined. In the models comparing shared and distinctive features trained in semantic blocks, there was a significant negative effect of feature type on response time blocks (t=-2.84, p=.005), indicating that participants were faster to verify distinctive features than shared features. There was also a marginally significant negative interaction between feature type and session for response time (t=-1.72, p=.085), indicating that participants had greater increases in speed (decreases in response time) for the verification of distinctive features relative to shared features. These effects on response time did not correspond to significant effects on accuracy (z=0.79, p=.429 for the main effect of feature type; z=1.13, p=.261 for the interaction of feature type and session), although these effects were numerically in the expected direction. In the models comparing shared features trained in semantic blocks to distinctive features trained in other contexts there was a significant advantage for distinctive features in terms of both accuracy (z=4.23, p<.001) and response time (t=-4.97, p<.001). This advantage increased across sessions for both accuracy (z=3.27, p=.001) and response time (t=-2.57, p=.057). These results were consistent with the hypothesis: training in semantic blocks results in an advantage for distinctive features relative to shared features, indicating increased distinctiveness.

Other significant effects were also observed in this analysis. Throughout the written semantic probe task, participants learned as evidenced by the main effects of session on accuracy (z=1.96, p=.050 for the model comparing shared and distinctive features trained in semantic blocks; z=2.86, p=.004 for the model comparing shared features trained in semantic blocks to distinctive features trained in other contexts) and response time in both analyses (t=-5.11, p<.001 for the model comparing shared and distinctive features trained in semantic blocks; t=-4.97, p<.001 for the model comparing shared features trained in semantic blocks to distinctive features trained in other contexts). There was a marginally significant effect of days since the last session on accuracy in the model comparing shared features trained in semantic blocks to distinctive features trained in other contexts (z=1.65, p=.099), suggesting that increased time between sessions benefitted performance to some extent, although this effect was not significant in any of the other models.

Spoken semantic probe results over sessions. Figure 17 shows the results of the spoken semantic probe task over sessions. Tables D7 and D8 (in Appendix D) report the results of the analyses of this task.

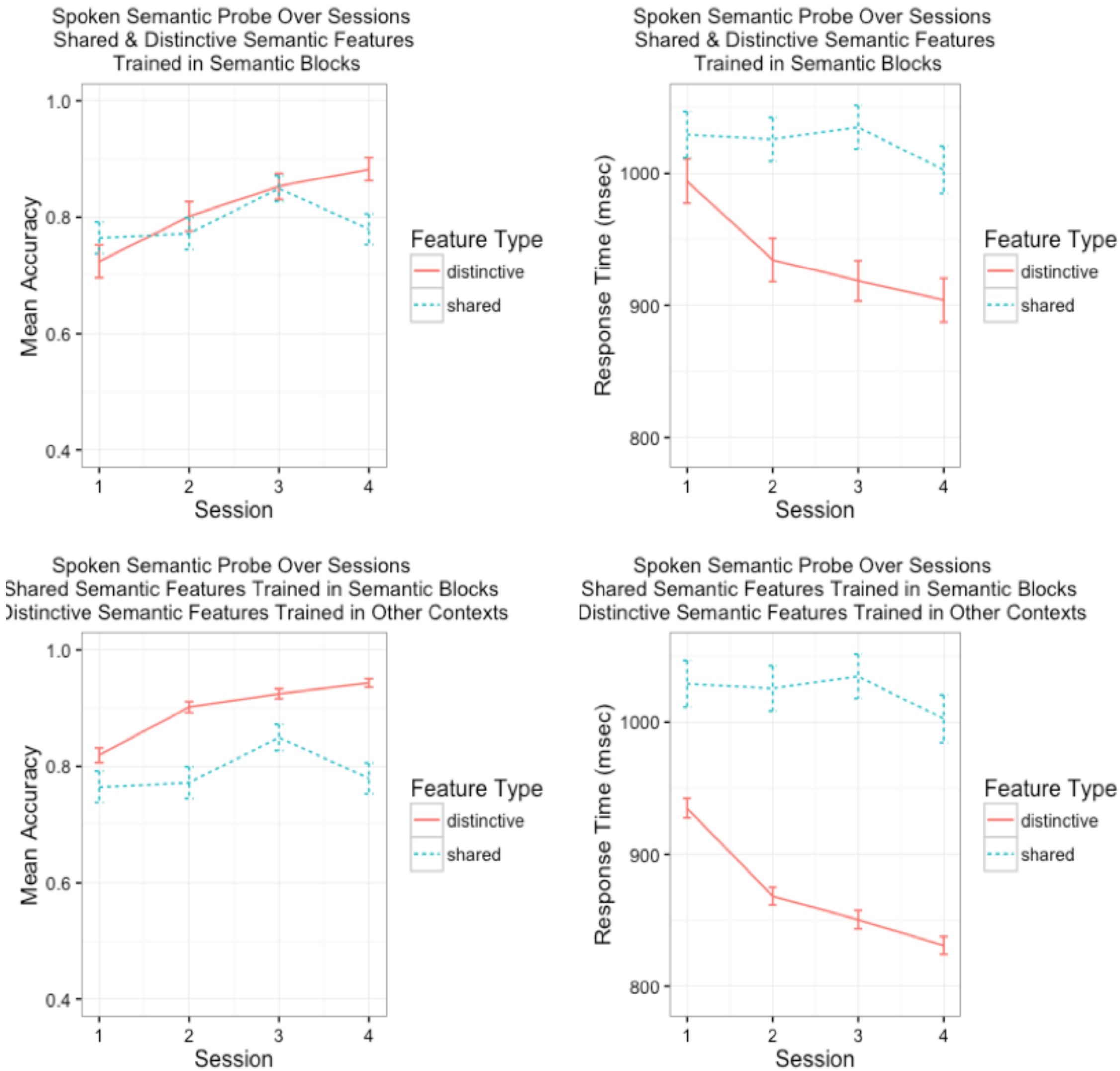


Figure 17. Results of the semantic probe task over sessions from the spoken experiment.

The top left panel shows accuracy for verification of semantic features across sessions, comparing shared and distinctive features trained in semantic blocks.

The top left right shows response time for verification of semantic features across sessions, comparing shared and distinctive features trained in semantic blocks.

The bottom left panel shows accuracy for verification of semantic features across sessions, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

The bottom left right shows response time for verification of semantic features across sessions, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

The results of the analysis of the semantic probe data over sessions for the spoken experiment were generally consistent with the written experiment. As in the written experiment, there was a main effect of feature type such that participants were faster to verify distinctive features trained in semantic blocks than shared features over the entire experiment (t=-6.01, p<.001), although this main effect was not significant for the accuracy analysis (z=1.54, p=.124). Significant interactions of feature type and session showed that the advantage for distinctive features increased across sessions in terms of both accuracy (z=2.95, p=.003) and response time (t=-2.24, p=.025). Participants were both faster (t=-6.32, p<.001) and more accurate (z=4.25, p<.001) to verify distinctive features trained in segmental and unrelated contexts than shared features trained in semantic contexts over the entire experiment. Again, this advantage for distinctive features grew as more training sessions were completed in terms of accuracy (z=4.01, p<.001) and response time (t=-3.49, p<.001). The consistent advantage for distinctive features relative to shared features suggests that training in semantic blocks increased distinctiveness as predicted according to the e-ILM.

In terms of other effects found in this analysis of semantic probe data from the spoken experiment, participants again demonstrated learning. They improved in both accuracy (z=2.78, p=.005 for the model comparing shared and distinctive features trained in semantic blocks; z=3.87, p<.001 for the model comparing shared features trained in semantic blocks to distinctive features trained in other contexts) and response time (t=-1.90, p=.058 for the model comparing shared and distinctive features trained in semantic blocks; t=-2.56, p=.010 for the model comparing shared features trained in semantic blocks to distinctive features trained in other contexts) across sessions. There were no significant effects of days since the last session, indicating that spacing training sessions further apart in time did not impact performance on this task.

Written and spoken semantic probe results over sessions. Taken together, the results of the written and spoken experiments are consistent: when examining the semantic verification probe task over sessions, there is an advantage for distinctive features relative to shared features. Numerically, this advantage appeared for all main effects of feature type and interactions of feature type and session in all eight models of the semantic probe data over sessions. This is in line with the hypothesis that training in semantic blocks strengthens distinctive features but weakens shared features, enhancing distinctiveness.

Segment probe results over sessions. The results of the analyses of the written and spoken segment probe tasks over the four training sessions are presented below.

Written segment probe results over sessions. Figure 18 shows the results of the written segment probe task (letter probe task) over sessions. Tables D9 and D10 (in Appendix D) report the results of the analyses of this task.

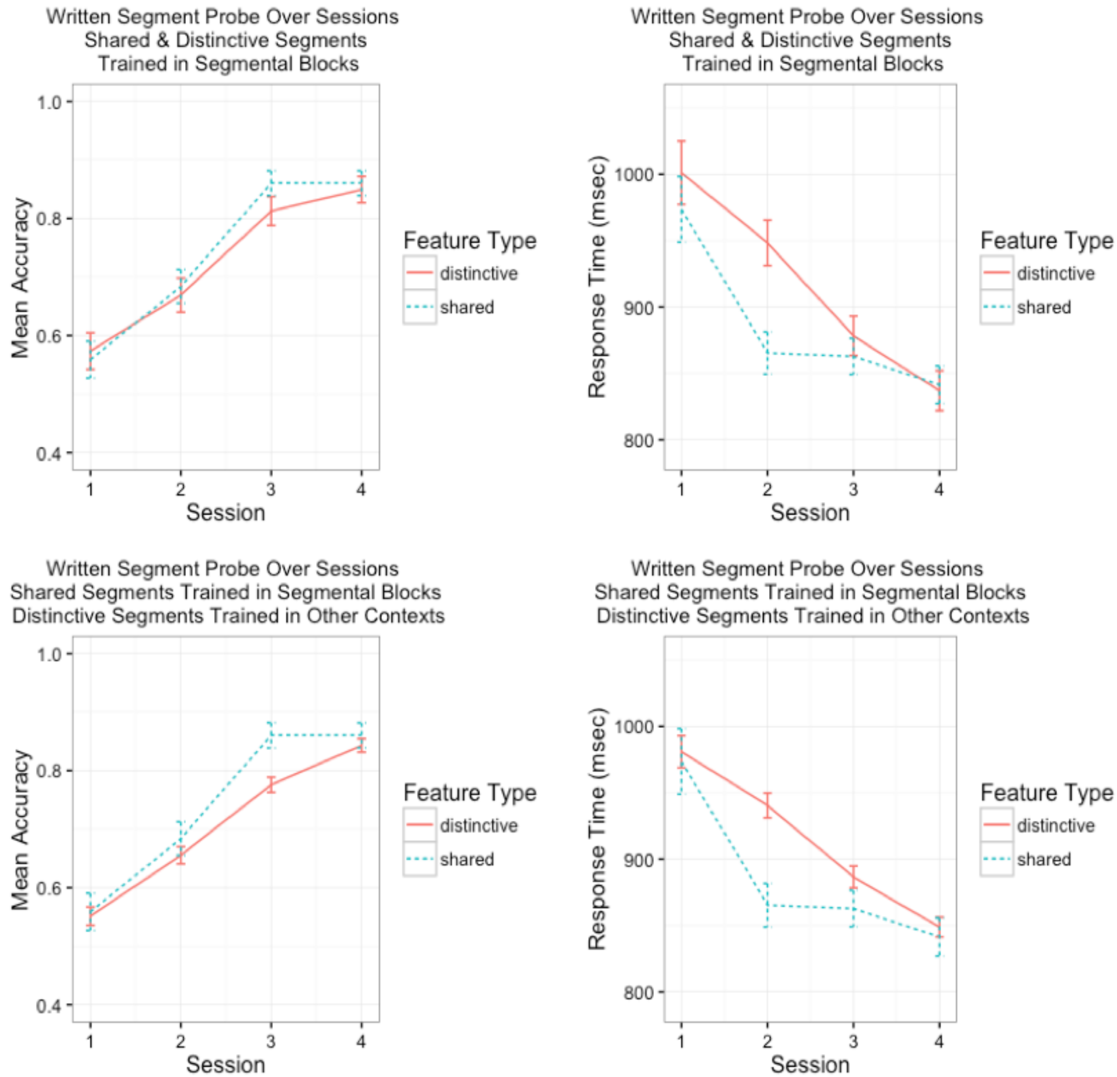


Figure 18. Results of the segment probe task over sessions from the written experiment.

The top left panel shows accuracy for verification of segments across sessions, comparing shared and distinctive segments trained in segmental blocks.

The top left right shows response time for verification of segments across sessions, comparing shared and distinctive segments trained in segmental blocks.

The bottom left panel shows accuracy for verification of segments across sessions, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

The bottom left right shows response time for verification of segments across sessions, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

According to the e-ILM predictions, training in segmental blocks should reduce distinctiveness. As in the analysis of the semantic probe data, in this analysis of written segment probe data the main effect of feature type and the two-way interaction of feature type and session were examined in order to evaluate the prediction. In the models comparing shared and distinctive segments trained in segmental blocks, there was a significant positive effect of feature type on response time (t=2.27, p=.023) and a marginally significant negative effect of feature type on accuracy (z=-1.78, p=.075), indicating that participants were faster and more accurate to verify shared segments relative to distinctive ones. This advantage for shared features increased as more training was completed, as indicated by the marginally significant negative interaction of feature type and session for accuracy (z=-1.72, p=.085). There was not a significant interaction for response time (t=-1.40, p=.160) although the effect was again numerically in the direction of an increasing advantage for shared features. In the models comparing shared segments trained in segmental blocks to distinctive features trained in other contexts, there were no significant main effects of feature type (z=-0.92, p=.357 for accuracy; t=1.25, p=.212 for response time) or interactions of feature type and session (z=-1.54, p=.123 for accuracy; t=-0.84, p=.401 for response time) although the effects were in the expected direction. The results of the models comparing shared and distinctive segments trained in segmental blocks fit the prediction of the e-ILM: training in segmental blocks resulted in an advantage for shared segments relative to distinctive segments, indicating reduced distinctiveness. This is the reverse direction of the effect found for the semantic probe task, in which training in a semantic block enhanced distinctiveness.

Looking at the other effects in the models showed that over the course of the written segment probe task, participants demonstrated learning that led to more accurate responses (z=6.88, p<.001 for the model comparing shared and distinctive segments trained in segmental blocks; z=7.68, p<.001 for the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts) and faster responses (t=-2.02, p=.044 for the model comparing shared and distinctive segments trained in segmental blocks; t=-1.76, p=.078 for the model comparing shared segments trained in segmental blocks to distinctive segments trained in other contexts) across sessions. There were no significant effects of days since the last session, indicating that differences in time between sessions did not influence performance here.

Spoken segment probe results over sessions. Figure 19 shows the results of the spoken segment probe task (sound probe task) over sessions. Tables D11 and D12 (in Appendix D) report the results of the analyses of this task.

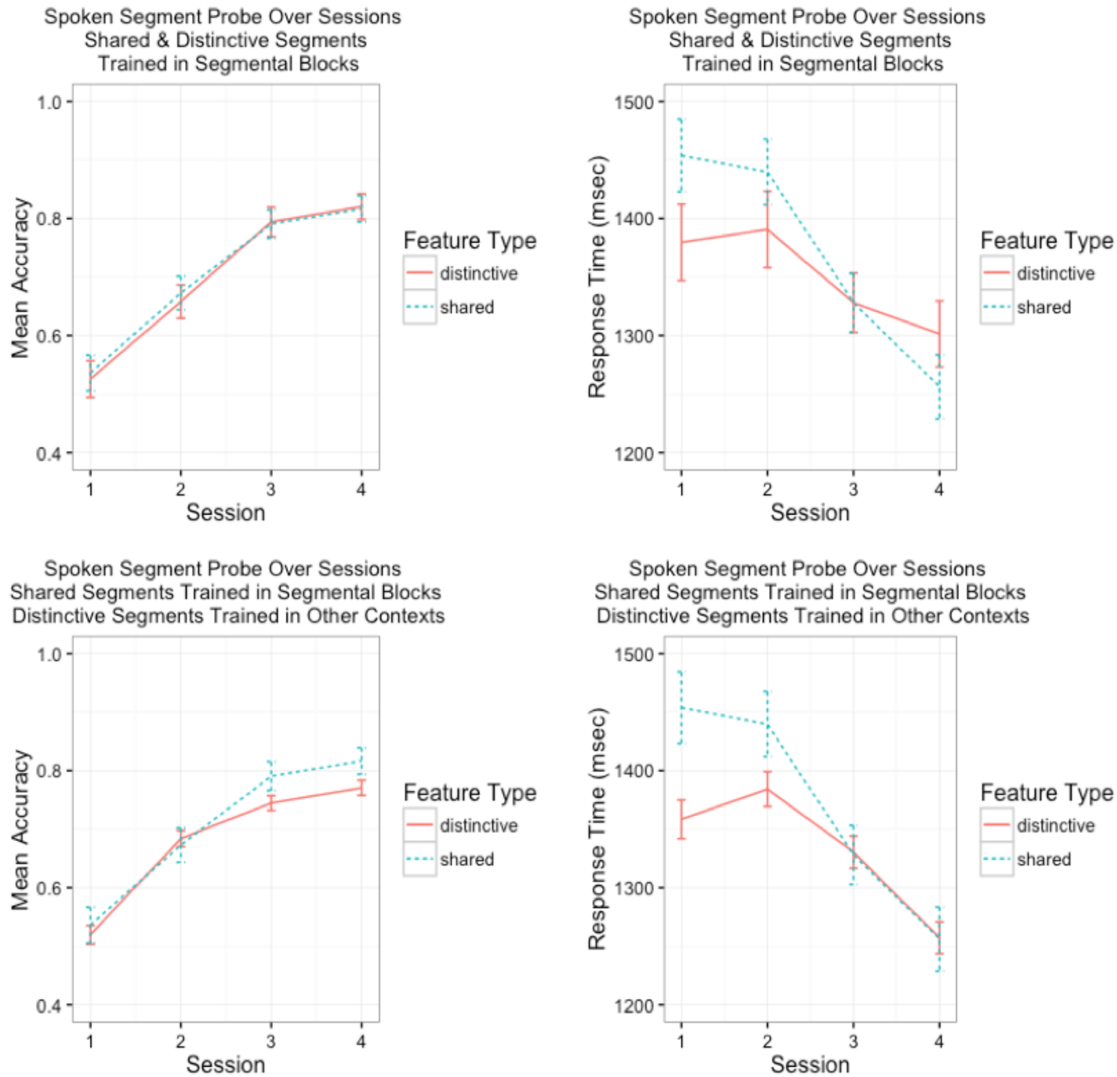


Figure 19. Results of the segment probe task over sessions from the spoken experiment.

The top left panel shows accuracy for verification of segments across sessions, comparing shared and distinctive segments trained in segmental blocks.

The top left right shows response time for verification of segments across sessions, comparing shared and distinctive segments trained in segmental blocks.

The bottom left panel shows accuracy for verification of segments across sessions, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

The bottom left right shows response time for verification of segments across sessions, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

In the analysis of the spoken segment probe task, there were no main effects of feature type on performance (z=-0.33, p=.743 for accuracy, t=-0.46, p=.644 for response time in the model comparing shared and distinctive segments trained in segmental blocks; z=-1.00, p=.316 for accuracy, t=-0.76, p=.448 for response time in the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts). However, there were consistent marginally significant interactions between feature type and session. In the model comparing shared and distinctive segments trained in segmental blocks, participants became faster to verify shared segments relative to distinctive ones as they completed more training session (t=1.92, p=.055) although there was not a significant impact on accuracy (z=0.04, p=.969). In the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts, participants became both more accurate (z=-1.79, p=.073) and faster (t=1.83, p=.067) to verify shared vs. distinctive segments over sessions. While these results are statistically weak, the increased advantage for shared segments is again consistent with the prediction that training in segmental blocks reduces distinctiveness.

Turning to the other effects in the analysis, as in the other analyses of probe tasks over sessions, there were significant effects of session on accuracy (z=6.80, p<.001 for the model comparing shared and distinctive segments trained in segmental blocks; z=8.56, p<.001 for the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts) and response time (t=-2.79, p=.005 for the model comparing shared and distinctive segments trained in segmental blocks; t=-3.44, p=.001 for the model comparing shared segments trained in segmental blocks to distinctive segments trained in other contexts) as participants learned the items.

Participants were slower to respond when more days intervened between sessions (t=2.15, p=.032 for the model comparing shared and distinctive segments trained in segmental blocks; t=3.31, p=.001 for the model comparing shared segments trained in segmental blocks to distinctive segments trained in other contexts), although there were not corresponding effects on accuracy.

Written and spoken segment probe results over sessions. Together, the written and spoken segment probe task results are in line with the hypotheses of the e-ILM. They provide support for the prediction that training in segmental blocks reduces distinctiveness. Although the pattern of results was not identical across modalities, with significant main effects of feature type in the written experiment and interactions of feature type and session in the spoken experiment, both showed an advantage for shared segments relative to distinctive segments. This advantage appeared numerically in all eight models for 11/16 effects (counting main effects of feature type and interactions of feature type and session as separate effects). The advantage for shared segments relative to distinctive segments contrasts with the advantage for distinctive features relative to shared features found in the semantic feature probe tasks.

Distinctiveness at follow-up. Distinctiveness analyses next considered the results of the probe tasks collected at follow-up two weeks after the final training session.

Model structure. Similar to the analyses of distinctiveness changes across training reported above, the analyses of distinctiveness effects at follow-up relied on multilevel mixed models with random effects constructed using the same architecture. As in the previous analysis, there were sixteen total models, one for each combination of modality (written or spoken), task (semantic probe or segment probe), data type (accuracy or response time), and distinctiveness comparison (comparing shared features/segments to distinctive features/segments trained in the same or different blocks).

In the spoken experiment models, fixed effects included feature type (shared or distinctive) and days since the last session. Feature type was a categorical variable coded as shared=-1, distinctive=1. Days since the last session was treated as a continuous variable that was centered and scaled. It was not possible to include a full random effects structure due to failures of convergence, so random slopes over items were not included. The resulting random effects structure included random intercepts for subjects and items as well as random slopes for feature type and days since the last session over subjects.

Written experiment model structure was the same as spoken model structure except that the written experiment models did not include days since the last session as there was a limited distribution of days in the written experiment: 15/17 participants completed follow-up testing exactly 14 days after the final training session. In order to confirm that days should be included in the spoken models but not the written models, effects on model fit of including vs. excluding days since the last session were evaluated using -2 times the change in log likelihood, which is distributed as χ^2 with the number of parameters added equal to the degrees of freedom. Model comparisons showed that there were not significant differences in fit between written models that included and excluded days since the last session (all had $p>.49$). Excluding days was preferred as interpreting the effect of days in this situation would more accurately be described as interpreting individual differences for the two individuals who completed follow-up assessments 12 and 17 days, respectively, after the final training session as compared to the group. On the other hand, including days improved fit for several spoken models (for segment probe accuracy comparing shared segments and distinctive segments trained in segmental blocks, $\chi^2(4)=9.86$, $p<.05$; for segment probe accuracy comparing shared segments trained in segmental blocks to distinctive segments trained in other blocks, $\chi^2(4)=8.42$, $p<.10$; and for semantic probe response time comparing shared features trained in semantic blocks to distinctive features trained in other blocks, $\chi^2(4)=8.32$, $p<.10$). Therefore, this variable was included for all models of the spoken experiment data, but not for models of the written experiment data.

Explanation of included variables. In the analysis of distinctiveness at follow-up, the critical effect was the main effect of feature type. If training in a semantically related context increases distinctiveness in a way that persists to the follow-up period, participants should be faster and more accurate to accept distinctive vs. shared features. If training in a segmentally related context reduces distinctiveness in a way that persists to the follow-up period, participants should be slower and less accurate to accept distinctive vs. shared features.

Days since the last session, included only in the spoken models, evaluated whether there is an effect of the spacing between the last training session and follow-up session on speed and accuracy of responses.

Semantic probe results at follow-up. The results of the analyses of the written and spoken semantic probe tasks at follow-up are presented below.

Written semantic probe results at follow-up. Figure 20 shows the results of the written semantic probe task at follow-up. Tables D13 and D14 (in Appendix D) report the results of the analyses of this task.

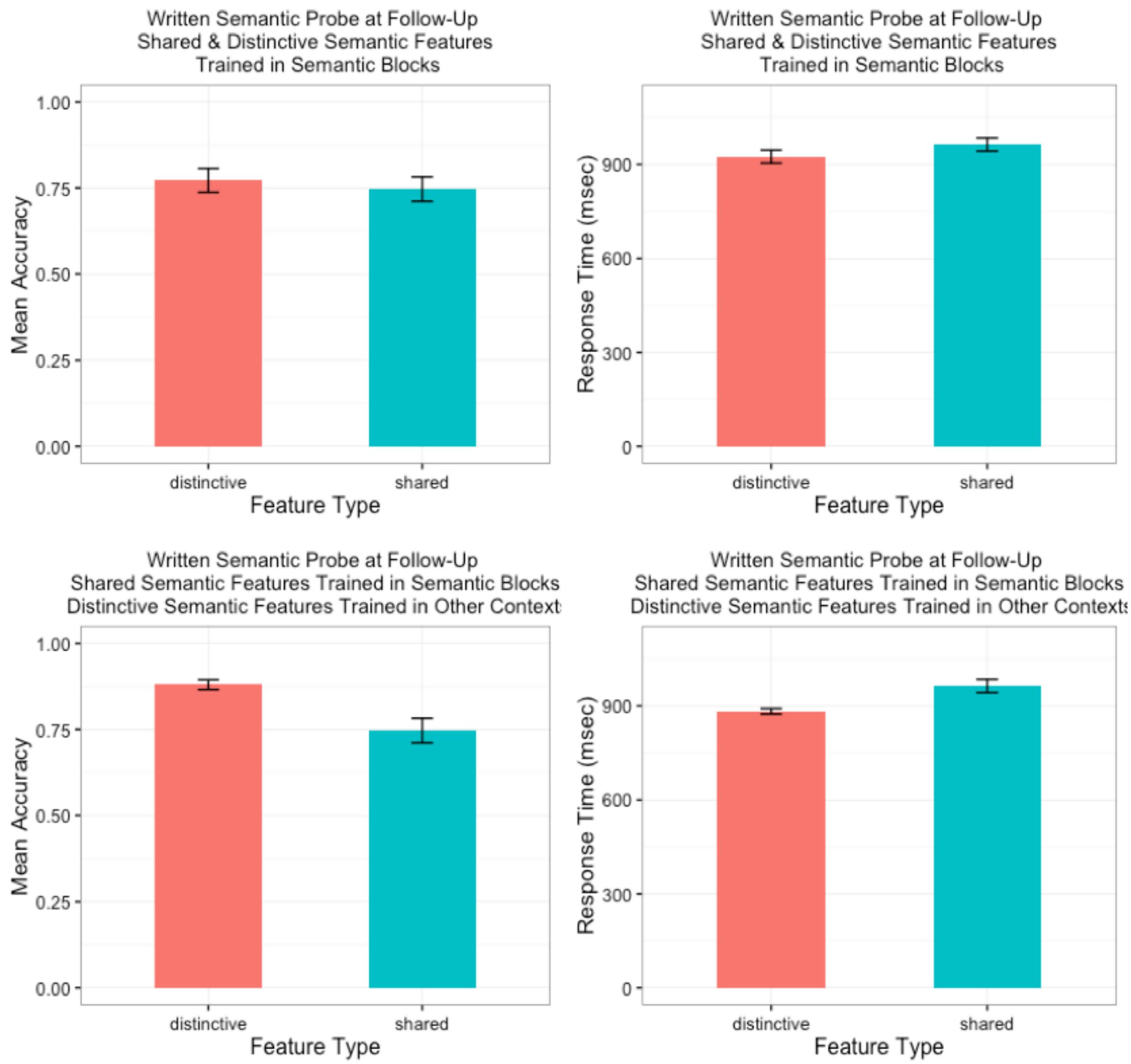


Figure 20. Results of the semantic probe task at follow-up from the written experiment.

The top left panel shows accuracy for verification of semantic features at follow-up, comparing shared and distinctive features trained in semantic blocks.

The top left right shows response time for verification of semantic features at follow-up, comparing shared and distinctive features trained in semantic blocks.

The bottom left panel shows accuracy for verification of semantic features at follow-up, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

The bottom left right shows response time for verification of semantic features at follow-up, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean.

According to the e-ILM predictions, training in semantic blocks should increase distinctiveness. To evaluate this prediction, the analyses of distinctiveness at follow-up included the main effect of feature type. For the written semantic probe task, in the models comparing shared and distinctive features trained in semantic blocks, there was a marginally significant advantage for distinctive features in terms of response time (t=-1.87, p=.061) but not accuracy (z=0.52, p=.601) although the numerical effect was in the expected direction. In the models comparing shared features trained in semantic blocks to distinctive features trained in other contexts, there was a significant advantage for distinctive features in terms of both accuracy (z=4.01, p<.001) and response time (t=-3.90, p<.001). Just as in the analysis of distinctiveness over sessions, these results are in line with the hypothesis that semantic blocking increases distinctiveness.

Spoken semantic probe results at follow-up. Figure 21 shows the results of the spoken semantic probe task at follow-up. Tables D15 and D16 (in Appendix D) report the results of the analyses of this task.

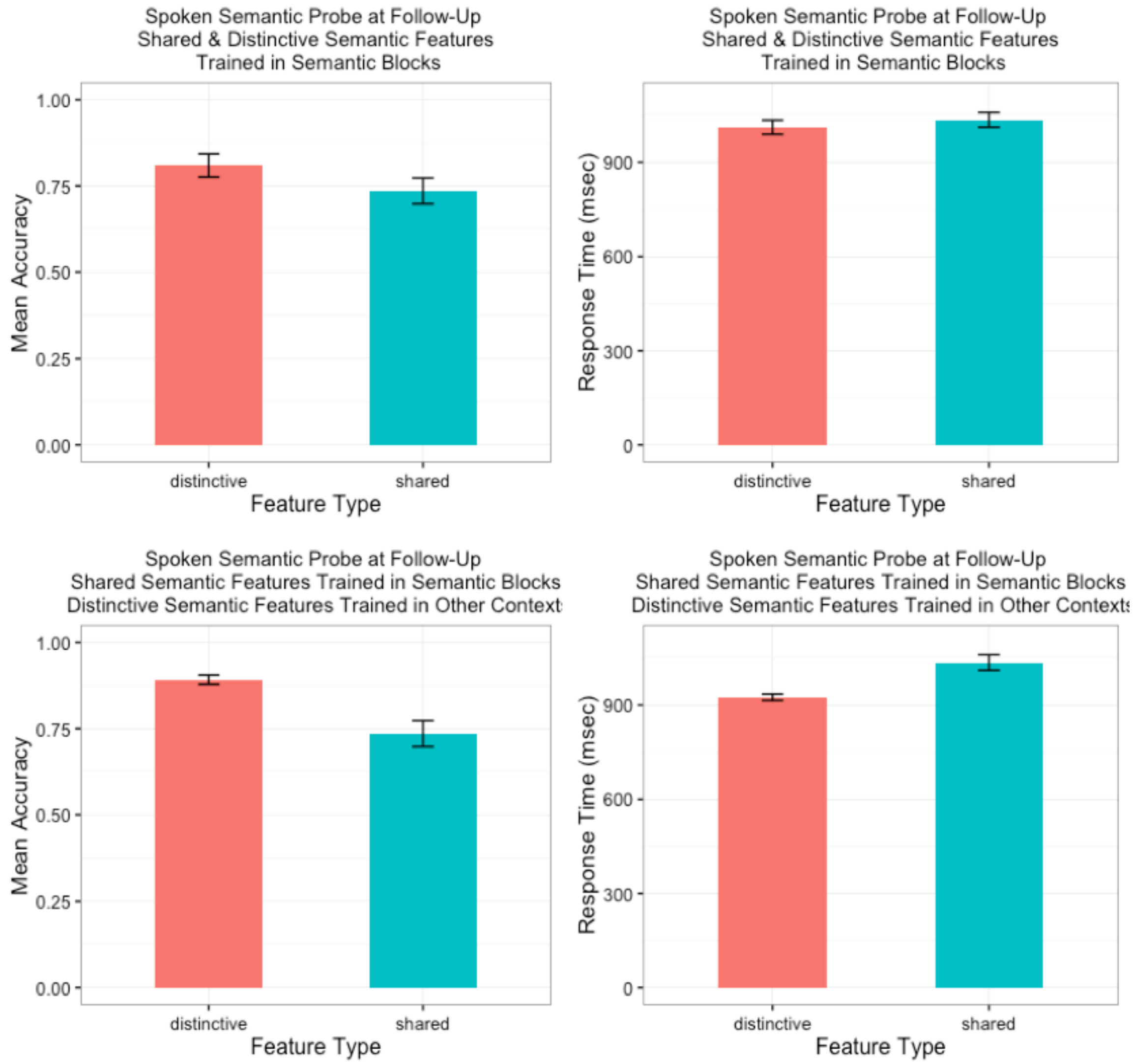


Figure 21. Results of the semantic probe task at follow-up from the spoken experiment.

The top left panel shows accuracy for verification of semantic features at follow-up, comparing shared and distinctive features trained in semantic blocks.

The top left right shows response time for verification of semantic features at follow-up, comparing shared and distinctive features trained in semantic blocks.

The bottom left panel shows accuracy for verification of semantic features at follow-up, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

The bottom left right shows response time for verification of semantic features at follow-up, comparing shared features trained in semantic blocks to distinctive features trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean.

For the spoken semantic probe task analysis of distinctiveness at follow-up, there was a marginally significant advantage for distinctive features trained in semantic blocks over shared features trained in semantic blocks in terms of accuracy (z=1.86, p=.062) but not in terms of response time (t=-1.14, p=.252) although the numerical effect was in the expected direction. As in the written experiment and the analysis of the probe task over sessions, there was a significant advantage for distinctive features trained in other contexts in terms of both accuracy (z=4.72, p<.001) and response time (t=-3.60, p<.001).

Spoken experiment models also included the main effect of days since the last session. There was a significant effect of this variable on response time for the model comparing distinctive and shared features trained in semantic blocks (t=-2.67, p=.008), indicating that participants who experienced longer retention periods were faster to respond on the follow-up task. There were not significant effects of days since the last session in any of the other models.

Written and spoken semantic probe results at follow-up. The results of analyses of the semantic probe task in both the written and spoken modalities concur: training in semantic blocks leads to an advantage for distinctive features relative to shared features. This advantage was numerically present for all effects in all eight models. As predicted by the e-ILM, training in semantic blocks increases distinctiveness.

Segment probe results at follow-up. The results of the analyses of the written and spoken semantic probe tasks at follow-up are presented below.

Written segment probe results at follow-up. Figure 22 shows the results of the written segment probe task at follow-up. Tables D17 and D18 (in Appendix D) report the results of the analyses of this task.

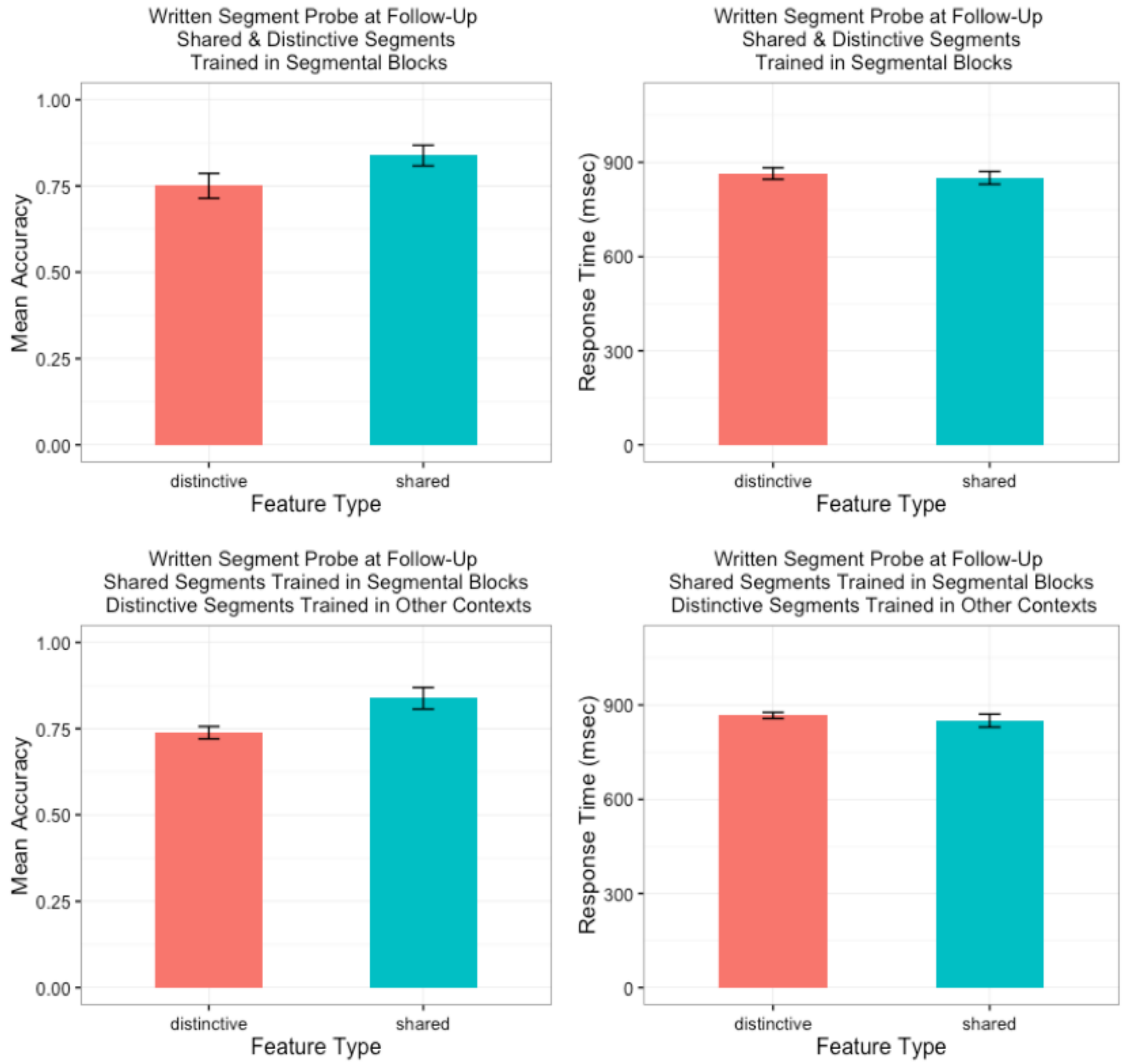


Figure 22. Results of the segment probe task at follow-up from the written experiment.

The top left panel shows accuracy for verification of segments at follow-up, comparing shared and distinctive segments trained in segmental blocks.

The top left right shows response time for verification of segments at follow-up, comparing shared and distinctive segments trained in segmental blocks.

The bottom left panel shows accuracy for verification of segments at follow-up, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

The bottom left right shows response time for verification of segments at follow-up, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean.

The e-ILM predicts that training in segmental blocks should reduce distinctiveness. To test this prediction, the main effect of feature type (shared or distinctive) was included in the analyses of distinctiveness at follow-up. For the written segment probe analysis at follow-up, participants were more accurate in responding to shared segments as opposed to distinctive segments (z=-2.81, p=.005 for the model comparing shared and distinctive segments trained in segmental blocks; z=-1.69, p=.092 for the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts). There were not significant effects of feature type on response time (t=1.27, p=.202 for the model comparing shared and distinctive segments trained in segmental blocks; t=1.03, p=.301 for the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts) although numerical trends were in the predicted direction. These results are in accord with the hypothesis that training in segmental blocks reduces distinctiveness.

Spoken segment probe results at follow-up. Figure 23 shows the results of the spoken segment probe task at follow-up. Tables D19 and D20 (in Appendix D) report the results of the analyses of this task.

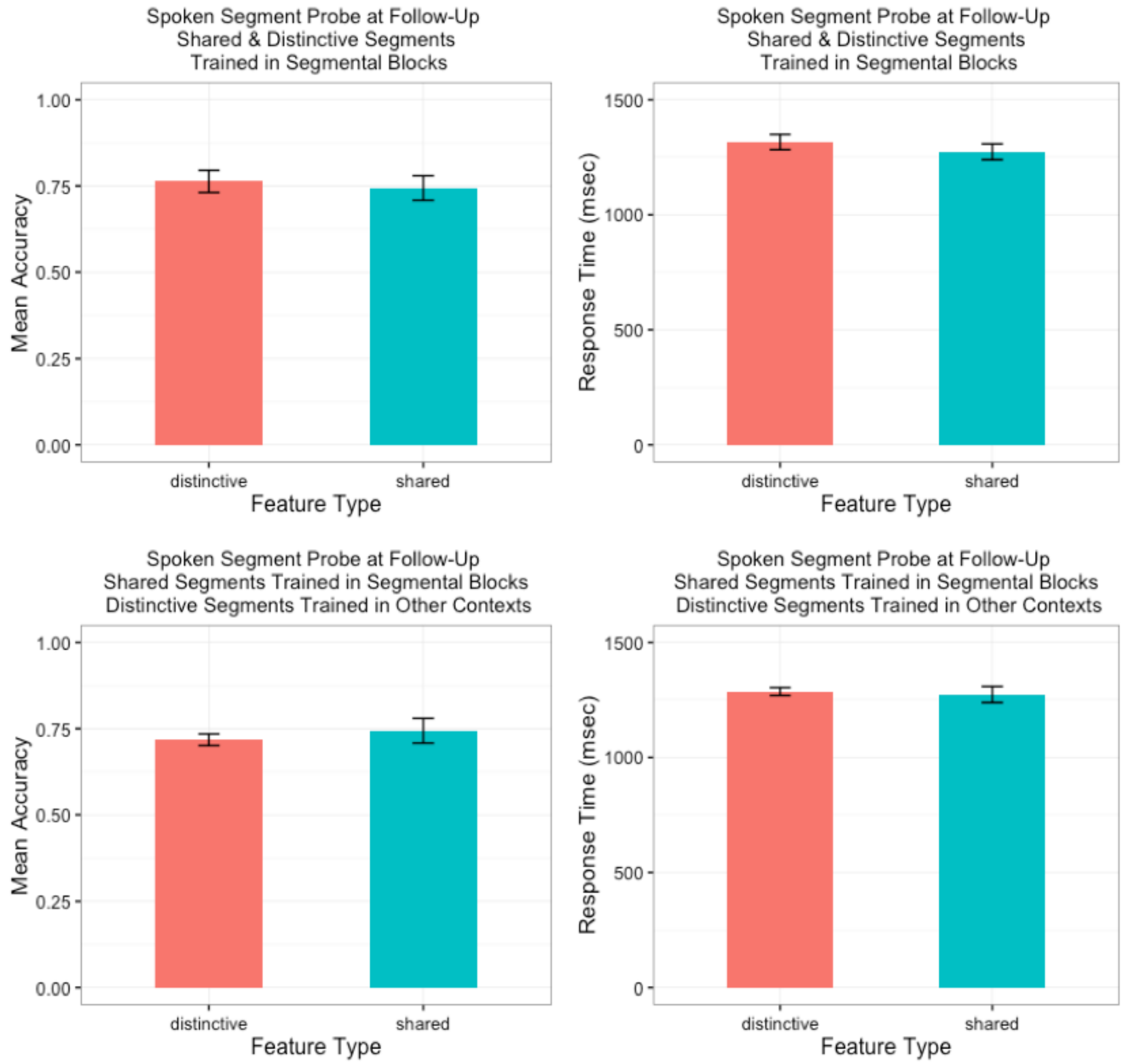


Figure 23. Results of the segment probe task at follow-up from the spoken experiment.

The top left panel shows accuracy for verification of segments at follow-up, comparing shared and distinctive segments trained in segmental blocks.

The top left right shows response time for verification of segments at follow-up, comparing shared and distinctive segments trained in segmental blocks.

The bottom left panel shows accuracy for verification of segments at follow-up, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

The bottom left right shows response time for verification of segments at follow-up, comparing shared segments trained in semantic blocks to distinctive segments trained in other contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean.

In the spoken segment probe analysis at follow-up, there were no significant effects of feature type for accuracy (z=0.84, p=.400 for the model comparing shared and distinctive segments trained in segmental blocks; z=-0.36, p=.723 for the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts) or response time (t=1.05, p=.293 for the model comparing shared and distinctive segments trained in segmental blocks; t=0.51, p=.607 for the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts). This does not clearly support the hypothesis that training in segmental blocks reduces distinctiveness of representations that persists to the follow-up period. However, note that three of the four models had effects that were numerically in the predicted direction.

There were significant effects of days since the last session on accuracy such that participants with longer retention intervals were more accurate (z=3.27, p=.001 for the model comparing shared and distinctive segments trained in segmental blocks; z=2.51, p=.012 for the model comparing shared segments trained in segmental blocks to distinctive features trained in other contexts). There were not significant effects of days since the last session on response time.

Written and spoken segment probe results at follow-up. The results of the written segment probe analysis provided support for the prediction that training in segmental blocks reduces distinctiveness: participants were less accurate in verifying distinctive segments than at verifying shared segments after a retention period of approximately two weeks. This effect appeared numerically in 4/4 models. However, in the spoken experiment, although there were numerical trends in 3/4 models, there were not significant differences between shared and distinctive features at follow-up, weakening support for the prediction.

Discussion of distinctiveness analyses. The analyses presented in this section evaluated the e-ILM predictions regarding the impact of training in blocked contexts on the distinctiveness of representations. Semantic blocking during training was predicted to increase distinctiveness, which should be observable as an advantage for distinctive features relative to shared features, while segmental blocking during training was predicted to decrease distinctiveness, which should be observable as an advantage for shared features relative to distinctive features.

The analyses of distinctiveness as measured by the semantic probe task were in line with the hypothesis that training in semantic blocks enhances distinctiveness. The analyses of the semantic probe task across sessions and at follow-up showed an advantage for distinctive features relative to shared features, suggesting that training in semantic blocks weakens shared features relative to distinctive features. This occurred in both the written and spoken experiments, which is unsurprising: although the training modality and names of trained items differed, the semantic probe tasks were identical in the written and spoken experiments.

In general, the analyses of distinctiveness as measured by the segment probe task were also consistent with the predictions. There was evidence of an advantage for shared segments relative to distinctive segments, indicating reduced distinctiveness as a result of training in segmental blocks. This advantage for shared segments was found across both the written and spoken experiments for the analysis over sessions, and it persisted over the retention period to reveal an accuracy advantage for shared segments at follow-up in the written experiment. For the spoken experiment, the results were statistically weaker: a number of the effects that support the predictions were marginally significant, and there were not significant effects of feature type at follow-up. Taken together, however, the results for the written and spoken experiments are consistent with the hypothesis of the incremental learning model.

Overall, the results of the distinctiveness analysis align with the hypotheses under investigation: semantic blocking increased distinctiveness, while segmental blocking reduced distinctiveness.

Retention

The last two analyses showed that training in both semantic and segmental blocks led to interference during acquisition relative to training in an unrelated context and that semantic blocking increased distinctiveness while segmental blocking reduced distinctiveness, confirming the direct predictions of the e-ILM. It is now possible to address the other e-ILM predictions that relate the effects of training to long-term learning. The third set of predictions relies on the assumption that increased distinctiveness leads to better long-term outcomes, while reduced distinctiveness conversely leads to worse long-term learning outcomes. If this assumption holds, the prediction states that training in semantic blocks should be beneficial for retention of names relative to training in an unrelated context. This is because training in semantic blocks should increase distinctiveness, a prediction that was supported by the previous analysis. On the other hand, the prediction regarding training in segmental blocks is that this type of blocking should be detrimental for retention of names relative to training in an unrelated context. This is because training in segmental blocks should reduce distinctiveness, another prediction that was supported by the previous analysis. The present analysis evaluates these predictions, characterizing the effects that training items in blocks has on the outcomes of learning as assessed by the recall tasks administered at the beginning and end of training sessions as well as at follow-up. If the predictions are accurate, recall should be better in terms of accuracy and/or response time for items trained in semantic blocks as compared to unrelated blocks, while recall should be worse in terms of accuracy and/or response time for items trained in segmental blocks as compared to unrelated blocks.

Overall performance on recall tasks. Overall, participants correctly produced the whole response on 75.2% of recall attempts in the written experiment and on 65.6% of recall attempts in the spoken experiment. The majority of errors were omissions in which no segments were produced (57.7% of errors in the written experiment and 66.0% of errors in the spoken experiment). There were a few within block substitutions whereby another name from the same block was produced instead of the target (2.6% of errors in the written experiment and 1.0% of errors in the spoken experiment) and across block substitutions whereby another name from the experiment that was not trained in the same block was produced instead of the target (12.1% of errors in the written experiment and 6.4% of errors in the spoken experiment). Remaining errors consisted of additions, deletions, and substitutions of segments (27.6% of errors in the written experiment and 26.6% of errors in the spoken experiment). As in the training analysis, further recall accuracy analyses examined whole response accuracy, not segment accuracy, a reasonable choice given that only 6.8% of trials in the written experiment and 9.1% of trials in the spoken experiment were not correct responses, whole omissions, or whole response substitutions.

Retention analysis structure: General. The analysis of data from the recall task separately considered two time scales for retention. First, retention over the few days between training sessions was assessed by comparing performance on the recall task administered at the end of one session and compared to performance on the recall task administered at the beginning of the next session. Second, retention over a longer period was assessed by examining performance at follow-up after a two-week period in which trained items were not practiced. This is a longer-term learning outcome measure. That is, similar to the distinctiveness analyses, retention was evaluated over sessions and at follow-up.

As in the previous analyses, the written and spoken experiments were modeled separately, as were accuracy and response time. As mentioned above, whole response accuracy was used. Response time analyses excluded incorrect responses as well as outliers greater than 2.5 standard deviations from each participant's mean. In order to evaluate the predictions of interest, semantic models compared items trained in semantic blocks to items trained in unrelated blocks while segmental models compared items trained in segmental blocks to items trained in unrelated blocks. This resulted in a total of eight models for the analysis over each time scale.

Recall across sessions. Retention analyses first considered the results of the recall task across the four training sessions.

Model structure. Recall accuracy and response time over the training sessions were analyzed using multilevel mixed models with random effects. Instead of dividing the experiment into the four sessions administered on separate days as in the training analysis, this analysis divided the experiment into three recall episodes defined by the retention period between sessions. Each recall episode included data from the recall task administered at the end of one session and from the recall task administered at the beginning of the next session. That is, recall episode 1 consisted of the recall tasks administered at the end of the first training session and the beginning of the second training session; recall episode 2 consisted of the recall tasks administered at the end of the second training session and the beginning of the third training session; and recall episode number 3 consisted of the recall tasks administered at the end of the third training session and the beginning of the fourth training session. Instead of looking at attempt within training session as was done in the training analysis, this analysis looked at relative time within recall episode. For each recall episode, the first administration of the recall task (given at the end of a training session prior to the retention period before the next training session) is referred to as the earlier relative time within recall episode, and the second administration of the recall task (given at the beginning of a training session after the retention period since the last training session) is referred to as the later relative time within recall episode. Models thus included the following fixed effects: block type (semantic, segmental, or unrelated context), relative time within recall episode (earlier or later), recall episode number (1-3), and the two- and three-way interactions between them. In this situation, days since the last recall attempt was not included as a variable since it was highly collinear with the relative time within recall episode (days for earlier was always 0 and days for later was always greater than 0). Due to lack of convergence, a full random structure could not be implemented in each model and the random slopes over items were dropped. This resulted in a model with random intercepts for subjects and items and a full random slope structure matching the fixed effect structure (block type, relative time within recall episode, recall episode number, and two- and three-way interactions between those) over subjects.

The variable of recall episode was centered and scaled. As in the previous analyses, block type was contrast coded as 1 and -1 in the models that compared each blocked context to the unrelated context. Relative time within recall episode was also contrast coded as earlier=-1, later=1.

Explanation of included variables. The choice to include relative time within recall episode as well as recall episode instead of including recall attempt across all training sessions was parallel to the choice made in the training analysis to include training attempt within session and session: it allows focus on the process of interest, retention. That is, it reduces the confound that training that occurs between recall attempts in the same session, which should increase performance within training sessions if training is effective, while performance is unlikely to increase and may in fact decrease as participants attempt to recall information retained across training sessions (recall episodes).

The main effect of recall episode shows effects of training over time, which are predicted to be positive for accuracy and negative for response time if training effectively enables participants to learn the new names.

The main effect of relative time within recall episode shows effects of recall over a retention period without training. A null effect would show that information is maintained over the retention period, while a negative effect on accuracy or a positive effect on response time would show that some information is forgotten over the retention period. A positive effect, although unlikely, would show that performance improves from the end of one training session to the beginning of the following training session.

The interaction between relative time within recall episode and recall episode tests whether the recall over the retention period is increased or reduced as more training is completed.

The main question of interest is whether recall is affected by the context of training, which was assessed through the two-way interaction of block type and relative time within recall episode. A positive interaction for accuracy or a negative interaction for response time would indicate that blocking increases retention, while the opposite interaction would indicate that blocking decreases retention. The hypothesis that semantic blocking is beneficial for retention predicts a positive interaction for accuracy and a negative interaction for response time in the semantic model comparing items trained in semantic blocks and items trained in an unrelated context. The hypothesis that segmental blocking is detrimental for retention predicts a negative interaction for accuracy and a positive interaction for response time in the segmental model comparing items trained in segmental blocks and items trained in an unrelated context.

The three-way interaction between block type, relative time within recall episode, and recall episode examines whether the effects of blocking on retention change as more training is completed.

Written recall results across sessions. Figure 24 shows the results of the recall task over sessions for the written experiment. Tables D21 and D22 (in Appendix D) report the results of the analyses of this task.

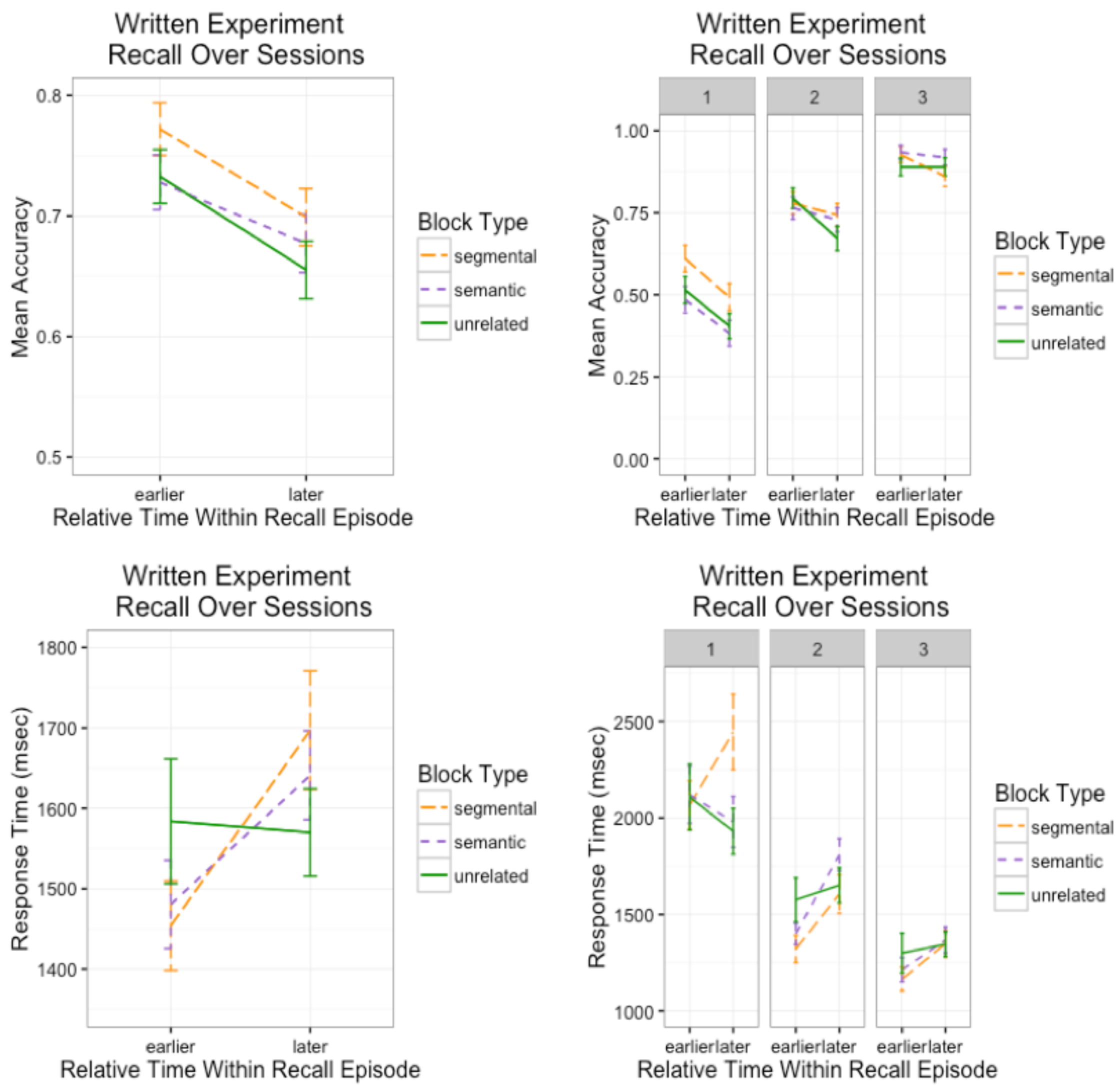


Figure 24. Results of the recall task over sessions from the written experiment.

The top left panel shows accuracy within recall episodes from the end of one training session to the beginning of the next, collapsed across the three recall episodes, for items trained in the three contexts.

The top right panel shows accuracy within the recall episodes from the end of one training session to the beginning of the next for each of the three recall episodes for items trained in the three contexts.

The bottom left panel shows response time within recall episodes from the end of one training session to the beginning of the next, collapsed across the three recall episodes, for items trained in the three contexts.

The bottom right panel shows response within the recall episodes from the end of one training session to the beginning of the next for each of the three recall episodes for items trained in the three contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

In the analysis of recall over the training sessions from the written experiment, the critical effect for evaluating the predictions about the impact of training in blocks on retention of item names was the two-way interaction of block type and relative time within recall episode. No significant interactions were found in these analyses (z=0.34, p=.737 for accuracy, t=0.27, p=.787 for response time for the semantic model; z=0.29, p=.773 for accuracy, t=0.91, p=.364 for response time for the segmental model). These results do not conclusively support any differences in retention from the end of one training session to the beginning of the next for items trained in blocked as opposed to unrelated contexts. This is not consistent with the predictions that training in semantic blocks should increase retention, while training in segmental blocks should decrease retention relative to training in unrelated contexts.

While the critical interaction was not significant, there was some weaker evidence suggesting that segmental blocking might negatively impact retention relative to training in unrelated contexts. Although there were no main effects of block type, there was a marginally significant interaction between block type and recall episode in the segmental model (z=-1.92, p=.055). This interaction indicates that participants were marginally significantly less accurate in recalling items trained in segmental contexts than unrelated contexts as more training was completed. While the interaction of block type and recall episode is not the most direct measure of blocking's effect on retention, the significant interaction does suggest that training in segmental blocks may detrimental for recall since these items do not increase in accuracy as much as items trained in unrelated blocks as more training is completed. This interaction was not significant in the segmental model of response time, nor was it significant in the semantic models of accuracy or response time.

In terms of effects that did not depend on block type, there was evidence that participants did learn the names of the items. The main effects of recall episode showed that participants increased in accuracy (z=10.49, p<.001 for the semantic model; z=9.87, p<.001 for the segmental model) and decreased in response time (t=-8.67, p<.001 for the semantic model; t=-9.40, p<0.001 for the segmental model) as they completed more training. There was also significant loss of information from the end of one training session to the beginning of the next session. The main effect of relative time within recall episode showed that participants decreased in accuracy (z=-2.81, p=.005 for the semantic model; z=-3.22, p=.001 for the segmental model) and increased in response time (t=3.61, p<.001 for the semantic model; t=3.69, p<.001 for the segmental model) from the earlier to later time points within the recall episode that corresponded to the end of one training session and beginning of the next. There were no significant two-way interactions between relative time within recall episode and recall episode, indicating that the recall over the retention period between training sessions neither increased nor decreased as participants completed more training sessions. Furthermore, there were no significant three-way interactions between block type, relative time within recall episode, and recall episode, showing that recall over the retention period did not change as more training sessions were completed based on block type.

Spoken recall results across sessions. Figure 25 shows the results of the recall task over sessions for the spoken experiment. Tables D23 and D24 (in Appendix D) report the results of the analyses of this task.

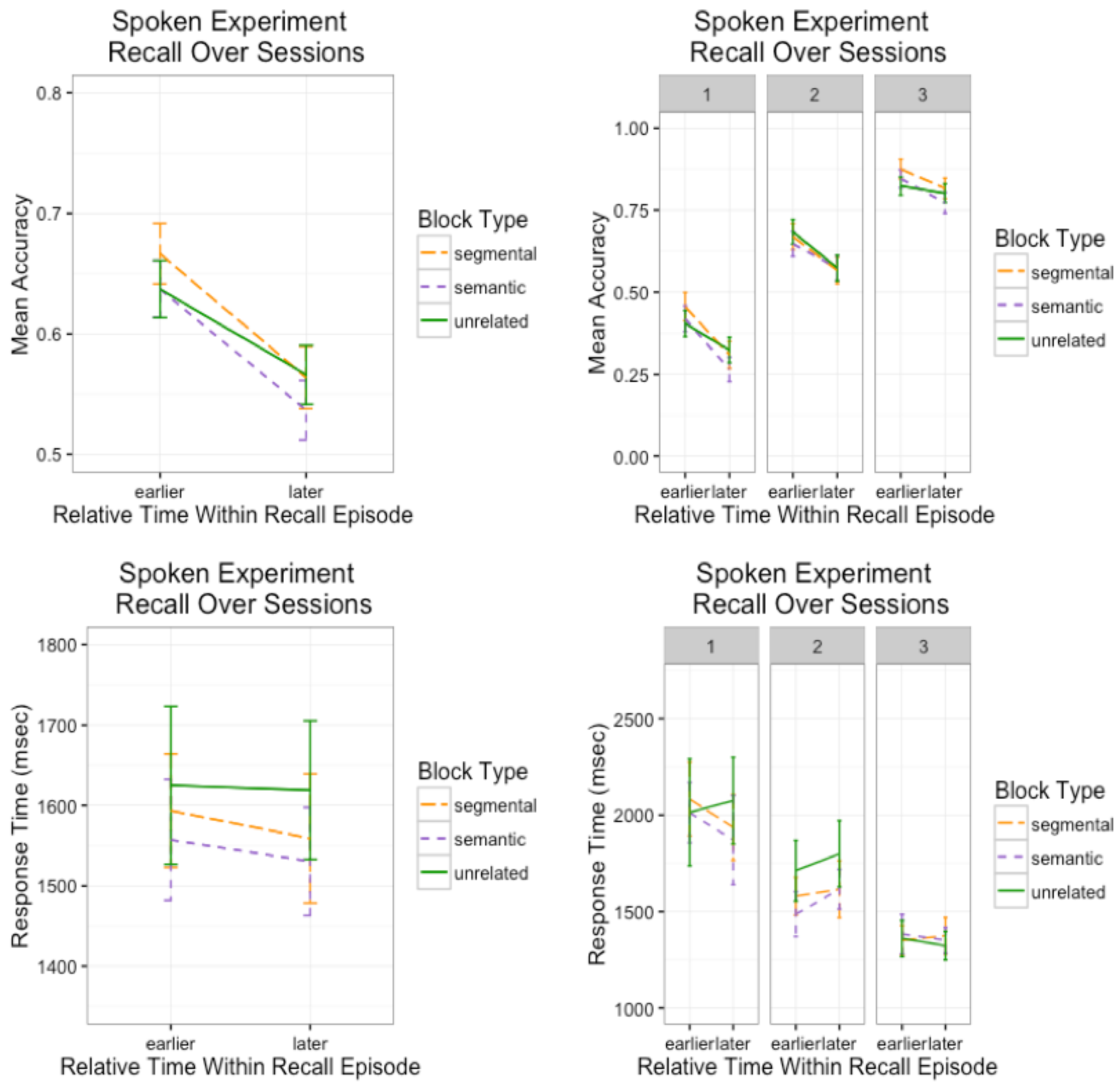


Figure 25. *Results of the recall task over sessions from the spoken experiment.*

The top left panel shows accuracy within recall episodes from the end of one training session to the beginning of the next, collapsed across the three recall episodes, for items trained in the three contexts.

The top right panel shows accuracy within the recall episodes from the end of one training session to the beginning of the next for each of the three recall episodes for items trained in the three contexts.

The bottom left panel shows response time within recall episodes from the end of one training session to the beginning of the next, collapsed across the three recall episodes, for items trained in the three contexts.

The bottom right panel shows response within the recall episodes from the end of one training session to the beginning of the next for each of the three recall episodes for items trained in the three contexts.

All figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

In the analysis of recall over the training sessions from the spoken experiment, the critical effect was again the two-way interaction of block type and relative time within recall episode. As in the written experiment, these interactions were not significant in the spoken experiment (z=-0.75, p=.453 for accuracy, t=0.30, p=.767 for response time for the semantic model; z=-0.47, p=.637 for accuracy, t=-0.57, p=.567 for response time for the segmental model). Again, these null effects do not support the hypotheses that block type affects retention of item names over sessions.

There were no other indications that training in blocks might impact retention. There were no main effects of block type or interactions between block type, relative time within recall episode, and/or recall episode. This indicates that there were no effects of training context on recall. Recall performance, including effects of learning across recall episodes and loss of information across the retention period between sessions, was the same regardless of whether items were trained in semantic, segmental, or unrelated contexts.

It is not the case that there were no significant effects in the analysis of spoken recall data across sessions. There was again evidence that participants learned the names. The main effects of recall episode showed both increases in accuracy (z=10.80, p<.001 for the semantic model; z=12.72, p<.001 for the segmental model) and decreases in response time (t=-6.79, p<.001 for the semantic model; t=-8.72, p<.001 for the segmental model) as participants completed more training. The main effect of relative time within recall session showed that participants decreased in accuracy (z=-3.58, p<.001 for the semantic model; z=-3.46, p=.001 for the segmental model) and increased in response time

(t=2.16, p=.031 for the semantic model) from the earlier to later time points within the recall episode for all analyses except the segmental model of response time (t=1.44, p=.149 for the segmental model). However, there were no significant two-way interactions between relative time within recall episode and recall episode, indicating that the recall over the retention period between training sessions did not change as participants completed more training sessions.

Written and spoken recall results across sessions. Overall, even though blocking increased difficulty during training and affected distinctiveness as predicted, there were not effects of training context on retention of names over training sessions for either modality when considering the interaction of block type and relative time within session, which directly focuses on recall across the retention period between sessions. While caution must be exercised in interpreting these null results, which are potentially indicative of low statistical power, the results of these analyses do not support the hypotheses of the e-ILM that predicted better recall for items trained in semantic blocks and worse recall for items trained in segmental blocks as compared to unrelated blocks. Numerical trends do not clarify the results: the numerical direction of the observed effects were consistent with a recall advantage for items trained in semantic vs. unrelated contexts in only 1/4 models and with a recall disadvantage for items trained in segmental vs. unrelated contexts in 2/4 models. Do note that there was some weak evidence in the written experiment that training in segmental vs. unrelated blocks is bad for recall as predicted. However, the marginally significant effect that supported this appeared in the interaction between block type and recall session. This interaction shows the combined effects of recall and training; it could be driven by the interference effect for segmental blocking that was observed during training as reported earlier in this chapter. There was no evidence of the predicted recall advantage for words trained in semantic blocks in either modality. One possible reason that the effects of blocking on retention were not seen in these analyses is the retention periods between sessions may not have been not ideal for observing long-term effects of training context. Retention over a longer time period of two weeks is considered in the next section.

Recall at follow-up. I next examined longer-term retention, looking at performance during the follow-up session two weeks after training was completed.

Model structure. As in the other models of recall across sessions presented above, recall accuracy and response time at follow-up was analyzed using multilevel mixed models with random effects. Models used to analyze the spoken experiment included block type (semantic, segmental, or unrelated context) and days since the last session as fixed effects. A full random structure was implemented in each model, with random intercepts for subjects and items, random slopes for block type and days since the last session over subjects, and random slopes for days since the last session over items.

As in the analyses of distinctiveness at follow-up, written models did not include days since the last session due to the restricted range of this variable in this experiment. Model comparisons showed that there were not significant differences in fit between written models that included and excluded days since the last session (all had $p>.90$), while including days improved fit for one spoken model (that of response time for the semantic vs. unrelated subset, $\chi^2(6)=12.65$, $p<.05$), so this variable was included for all models of the spoken experiment.

As in the models described previously, the continuous variable of days since the last session was centered and scaled in the spoken models. The categorical variable of block type was contrast coded as 1 and -1 in the models that compared each blocked context to the unrelated context.

Explanation of included variables. Here, the critical question of interest is whether recall at follow-up differs as a result of training in a blocked vs. unrelated context, so it is the main effect of block type that is of central interest[15]. According to the hypotheses presented above, the semantic blocking context should lead to better performance, while the segmental blocking context should lead to worse performance.

Days since the last session, included only in the spoken models, evaluated whether there is an effect of the length of the retention period between the last training session and follow-up session on speed and accuracy of responses.

Written recall results at follow-up. Figure 26 shows the results of the recall task at follow-up for the written experiment. Tables D25 and D26 (in Appendix D) report the results of the analyses of this task.

[15] An ideal analysis would also have taken into account performance at the end of training. However, the models that included this suffered from high collinearity and failures to converge and so are not reported. One limitation of the reported analysis is that this measure of recall at follow-up does not solely assess retention: it is also influenced by the effectiveness of training. It is possible that items may have worse recall performance at follow-up not because retention is worse for them but because training was less effective for them.

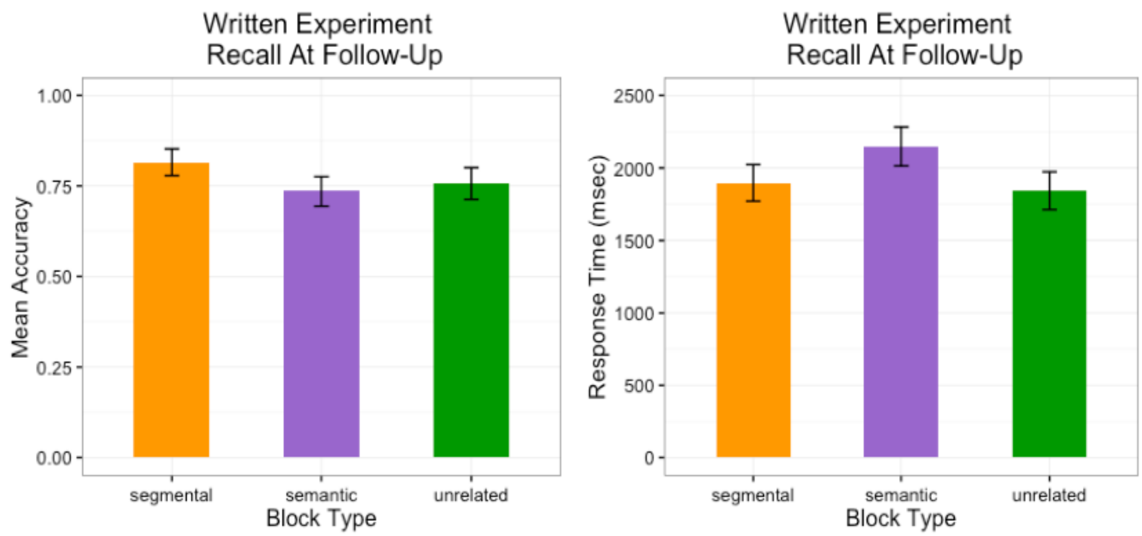


Figure 26. Results of the recall task at follow-up from the written experiment.

The left panel shows accuracy on the recall task at follow-up for items trained in the three contexts. The right panel shows response time for the recall task at follow-up for items trained in the three contexts. Both figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

In the analysis of the written recall task at follow-up, the critical question was whether recall differed at that point based on the context of training, which was assessed using the main effect of block type. There were no significant effects of block type on accuracy (z=0.13, p=.897 for the semantic model; z=1.29, p=.196 for the segmental model) or response time (t=1.43, p=.151 for the semantic model; t=-0.32, p=.749 for the segmental model) in the written experiment. These null results do not support the predictions of increased retention for items trained in semantic vs. unrelated blocks and of reduced retention for items trained in segmental vs. unrelated blocks. The null results observed here are consistent with the null interactions of block type and relative time within recall sessions reported above for the analysis of recall across sessions.

Spoken recall results at follow-up. Figure 27 shows the results of the recall task at follow-up for the spoken experiment. Tables D27 and D28 (in Appendix D) report the results of the analyses of this task.

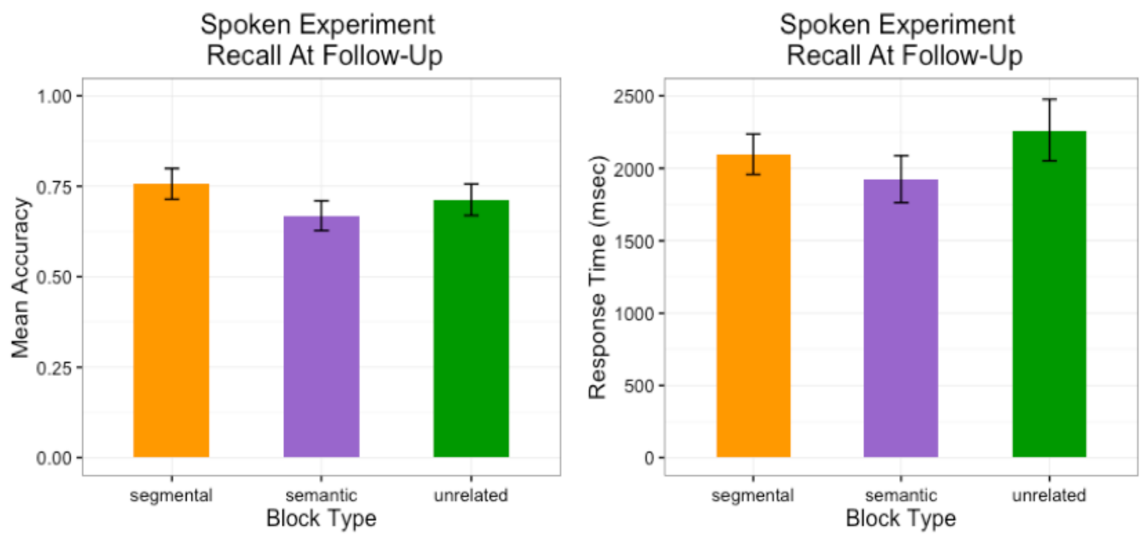


Figure 27. Results of the recall task at follow-up from the spoken experiment.

The left panel shows accuracy on the recall task at follow-up for items trained in the three contexts. The right panel shows response time for the recall task at follow-up for items trained in the three contexts. Both figures depict the mean of subject means. Error bars represent one standard error of the mean, corrected for repeated measures.

In the analysis of the spoken recall task at follow-up, the critical effect was again the main effect of block type. Here, there was a significant effect of block time on response time in the semantic model (t=-2.21, p=.027), indicating that participants were faster to produce items that had been trained in semantic blocks as opposed to unrelated contexts. This finding is consistent with the prediction that semantic blocking during training leads to improved retention as compared to training in an unrelated context. There was not a significant effect of block time on accuracy in the semantic model (z=-0.17, p=.869), nor was there a significant effect of block type on either accuracy (z=0.85, p=.393) or response time (t=-0.76, p=.445) in the segmental model. The null effects for segmental blocking do not support the prediction that segmental blocking leads to reduced retention as compared to training in an unrelated context.

No other significant effects were observed in this analysis; there were no significant effects of days since the last session.

Written and spoken recall results at follow-up. In the analysis of recall at follow-up, there was some weak support for the prediction that training in semantic blocks improves retention of names relative to training in unrelated blocks. While no significant effect of semantic vs. unrelated block type was observed for the written experiment (and a numerical trend was present in the expected direction for only accuracy not response time), participants were significantly faster to respond to items trained in semantic blocks as opposed to unrelated contexts at follow-up in the spoken experiment (note, however, that the numerical trend was opposite to the expected direction for accuracy). In terms of segmental vs. unrelated block type, there were no significant effects on retention of names at follow-up in either modality, and 0/4 models revealed numerical trends in the expected direction of reduced retention. These results do not support the prediction that training in segmental blocks reduces retention relative to training in unrelated blocks. However, it is again important to be careful in interpreting null effects, which may be due to low statistical power.

Discussion of retention analyses. Overall, the results of the retention analyses are less clear than the results of the training or distinctiveness analyses. Although the first two sets of e-ILM predictions held up to testing in this study—both semantic and segmental blocking led to interference during training relative to training in unrelated contexts, and semantic blocking increased distinctiveness while segmental blocking reduced distinctiveness—there was only limited support for the third set of predictions that semantic blocking during training should be detrimental for retention while segmental blocking during training should be beneficial for retention relative to training in unrelated contexts. In interpreting the results, it is important to keep in mind that the null effects observed for the majority of the retention analyses may be due to low statistical power; effect could potentially be observed in future studies with a larger number of items and/or participants.

Support for the prediction that semantic blocking positively impacts retention of names is drawn from the analysis of recall at follow-up in the spoken experiment: participants responded more quickly to items trained in semantic blocks than to items trained in unrelated contexts. This suggests that there is a long-term advantage for these items that persists over the two-week retention period between the final training session and the follow-up session. However, a similar advantage was not observed in the written version of the experiment, and there were not significant effects of semantic vs. unrelated blocking on retention across the shorter retention periods between the four training sessions.

There were no clear effects of segmental blocking during training on retention of names across training sessions or retention to follow-up. Results do not conclusively support increases or decreases in retention due to training in this type of block. The only suggestion that segmental blocking may negatively impact retention, in line with the prediction, came from an indirect analysis looking at the interaction of block type and recall episode in the analysis of recall across sessions in the written experiment. Participants showed marginally less improvement in accuracy as more sessions were completed for items trained in segmental vs. unrelated contexts. However, this effect is not a direct measure of recall, but instead may be affected by the interference observed during training. Ergo, there is minimal support for the prediction that segmental blocking is detrimental for retention.

Individual Differences: Correlations Between Training and Retention

The final set of e-ILM predictions considers the status of blocking as a desirable difficulty. Both semantic and segmental blocking were shown to increase interference during training (i.e., difficulty). According to a very broad construal of the learning literature, increasing difficulty during learning should lead to increased retention. However, the e-ILM provided a narrower focus to look at the specific effects of blocking, which may not be uniformly beneficial. According to this model, because of the way blocking during training changes distinctiveness, semantic blocking is predicted to increase retention while segmental blocking is predicted to reduce retention. That is, semantic blocking is predicted to be a desirable difficulty with opposite effects on training and retention, while segmental blocking is not.

Pattern of results across participants. One way to evaluate the predictions about desirable difficulty is to examine the pattern of results of training and retention analyses across participants. Across participants, semantic blocking led to interference during training relative to training in unrelated contexts. There was also some evidence consistent with a long-term advantage for items trained in semantic blocks vs. unrelated contexts: at follow-up, participants in the spoken experiment were faster to respond to items trained in semantic blocks than to items trained in unrelated contexts. This fits the general pattern of a desirable difficulty. Across participants, segmental blocking also led to interference during training relative to training in unrelated contexts. However, there was no direct evidence that this type of training led to either an advantage or a disadvantage for retention. Segmental blocking therefore does not fit the expected pattern of increased difficulty during training and reduced retention because of the inconclusive retention results.

Examining individual differences. Beyond analyzing the overall effects and patterns in the training and retention data, one can also more directly investigate the relationship between training and retention to evaluate whether either type of blocking is a desirable difficulty that increases retention to the extent that it increases difficulty during training. One way to accomplish this is to look at individual differences. Following Mirman (2014), the random effects for each subject can be extracted to give a measure of individual effect size that quantifies systematic deviations from the overall pattern predicted by the fixed effects. Correlations between the random effects for the same participants as they complete different tasks can be used to characterize the relationship between those tasks. For example, one can see if the participants with large effect sizes in one task are the same ones who have large effect sizes in another task.

Correlating blocking effects. In this analysis, I was interested in examining the relationship between the effects of blocking on training and on retention as a way to evaluate the predictions about desirable difficulties. The by-subject random effects of block type by training attempt within session (the critical effect in the training analysis) were extracted from the training analysis. Similarly, the by-subject random effects of block type by relative time within recall episode were extracted from the analysis of retention between training sessions, and the by-subject random effects of block type were extracted from the analysis of retention at follow-up (the critical effects in the retention analyses). Pearson's correlations between training and retention effects were then calculated. According to the hypotheses of the incremental learning model presented above, those participants who experience the most difficulty as a result of training in semantic blocks should show the biggest advantage for retention of items trained in semantic blocks. That is, a negative correlation is expected between the training and retention by-subject random effects for the models comparing semantic blocking to unrelated blocking. For segmental blocking, on the other hand, this should not be the case: segmental blocking should increase difficulty for both training and retention. If anything, a positive correlation would be expected between the training and retention by subject random effects for the models comparing segmental blocking to unrelated blocking. Alternatively, according to a more general desirable difficulties hypothesis whereby any increase in training difficulty leads to an increase in retention, there should be negative correlations between the by-subject training and retention effects for both the semantic vs. unrelated and segmental vs. unrelated models.

Results of individual differences analysis. Tables 5 and 6 report the results of the analyses of the correlations between training and recall effects.

Table 5. Correlations of by-subject effects of training and recall over sessions.

	Written				Spoken			
	Accuracy		RT		Accuracy		RT	
Correlation	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
block type (semantic vs. unrelated) * training attempt within session and block type (semantic vs. unrelated) * relative time within recall episode	-0.56	.020	-0.42	.093	0.04	.886	0.17	.513
block type (segmental vs. unrelated) * training attempt within session and block type (segmental vs. unrelated) * relative time within recall episode	-0.26	.309	0.19	.455	-0.04	.872	-0.34	.188

Table 6. Correlations of by-subject effects of training and recall at follow-up.

	Written				Spoken			
	Accuracy		RT		Accuracy		RT	
Correlation	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
block type (semantic vs. unrelated) * training attempt within session and block type (semantic vs. unrelated) at follow-up	0.10	.691	-0.38	.132	-0.02	.925	-0.19	.458
block type (segmental vs. unrelated) * training attempt within session and block type (segmental vs. unrelated) at follow-up	-0.15	.557	0.11	.668	-0.05	.845	-0.07	.781

First, the correlations between the by-subject random effects of block type by training attempt within session from the training analysis and the by-subject random effects of block type by relative time within recall episode from the recall over sessions analysis were examined. In the written experiment, there was a negative correlation for the models in which items trained in the semantic blocking context were compared to those trained in the unrelated blocking context, which was significant for accuracy (r=-.56, p=.020) and marginally significant for response time (r=-.42, p=.093).[16] This is in line with the hypothesis that semantic blocking is a desirable difficulty that increases interference during training but leads to better retention: the participants with the largest negative effects of semantic blocking on training were the same ones who had the largest positive effects on recall over sessions. Corresponding correlations were not significant in the spoken experiment. As in the retention analyses presented earlier, it is important not to overinterpret null effects: they may be due to low statistical power.

Turning to segmental blocking, correlations between training and recall over sessions effects was not significant for either accuracy or response time for the experiments in either modality. This suggests that there is not a clear relationship between training and recall difficulty for the segmental blocking context. This result is not consistent with the e-ILM prediction of a positive correlation between training and retention effects for items trained in segmental vs. unrelated blocks or with the more

[16] To confirm the significance of these results, Model-based Monte Carlo simulations were constructed to examine the correlations between random effects that would be expected by chance. Models were fit to simulated data 100 times, and the random effects were extracted and used to calculate correlations in the same way as was done for the real data. Results showed that the real correlation between training and retention effects on accuracy for the semantic vs. unrelated training contexts in the written experiment was more negative than all correlations from the simulations, suggesting that the real correlation was significant at the $p<.01$ level. The real correlation between training and retention effects on response time for the semantic vs. unrelated training contexts in the written experiment was more negative than all but 9/100 correlations from the simulations, suggesting that the real correlation was marginally significant at the $p<.09$ level.

general desirable difficulties prediction of a negative correlation. Again, the null effects may be due to low power.

The correlations between the by-subject random effects of block type by training attempt within session from the training analysis and the by-subject random effects of block type from the recall at follow-up analysis were also examined. However, there were no significant correlations for either the written or the spoken experiment.

Discussion of individual differences analyses. The results of the individual differences analysis add to the evidence from the pattern of results across participants when examining semantic blocking. There is some evidence that semantic blocking seems to be a desirable difficulty, although this evidence is relatively weaker than that supporting e-ILM predictions 1 and 2 regarding effects during training. While semantic blocking led to increased difficulty in training relative to training in unrelated contexts, it also led to increased retention to some extent. The individual differences analysis shows that there was a direct relationship between effects of training and retention over sessions in the written experiment, with the participants who experienced the most difficulty in training due to semantic blocking demonstrating the biggest retention advantage for those items over sessions. The pattern across participants had some results in the same direction: overall, participants experienced interference due to semantic vs. unrelated blocking in training, and in the spoken experiment they showed a retention advantage for those items relative to those rained in unrelated contexts in terms of response time at follow-up. Together, these results suggest that semantic blocking may be a desirable difficulty, in line with the e-ILM predictions.

There is also some agreement between the results of the individual differences analysis and the pattern of results across participants for segmental blocking. The individual differences analysis revealed no significant correlations between training and retention effects contrasting items trained in segmental blocks and unrelated contexts. Similarly, there was not a clear relationship in the pattern of results across subjects: although participants experienced interference due to segmental vs. unrelated blocking during training, there were no significant differences between retention for items trained in segmental blocks vs. those trained in unrelated contexts. These findings do not conclusively support the prediction of the e-ILM that the increased difficulty during training in segmental vs. unrelated contexts should lead to reduced long-term learning. While the specific predictions about how segmental blocking affects training were upheld (predictions 1 and 2), there was less agreement with the predictions that connect training effects to long-term learning outcomes (predictions 3 and 4).

General Discussion

The purpose of the study presented in this chapter was to test the predictions of the e-ILM (Table 4 in Chapter 3). There were different predictions regarding the effects of semantic and segmental blocking on the trajectory and outcomes of learning due to the proposed effects on the distinctiveness of representations.

In the present study, these predictions were tested as participants were trained and tested on the names and features of novel objects. Training occurred in blocks: items were trained in the context of other semantically related items that shared features, other segmentally related items that shared letters or sounds, or unrelated items. Participants were tested on production of the words at the beginning and end of every training session.

They also completed probe tasks at the end of every session in which they verified whether a given feature or segment was true for a given object. Comparisons of performance on distinctive and shared features were carried out to assess the effects of training on distinctiveness.

Separate experiments with different participants were conducted in the written and spoken modalities. In general, the pattern of results was consistent across modalities, especially for the training and distinctiveness analyses that evaluated the first two predictions concerning the direct extension of the model from production of known words to learning of new words. This suggests that similar principle underlie processing in both modalities. There were some differences between the results for the different modalities, such as whether effects appeared in accuracy or response time analyses, which may be due to differences in the modalities such as the time course of response execution.

Below, I consider how the results found in this study relate to the e-ILM predictions.

Semantic Blocking

First, consider the predictions regarding the effects of training in semantically related blocks.

Semantic prediction 1: Training in semantic blocks leads to interference during acquisition relative to training in unrelated contexts. This prediction was investigated by comparing the improvement for items trained in semantic and unrelated blocks as participants practiced producing each item more times within a training session. According to the prediction, participants should show reduced improvement in speed and/or accuracy during training for items in semantic blocks relative to items in unrelated blocks. Results showed that although participants successfully learned in both training contexts, the improvement within sessions for items trained in semantic blocks was slower than for items trained in unrelated contexts as measured by the interaction between block type and training attempt within session. There were differences across modality such that the interaction was significant in the analysis of accuracy in the written experiment and in the analysis of response time in the spoken experiment, which may be due to the different time courses of production whereby written production takes place over a longer period than does spoken production. However, the overall pattern of results was consistent with the prediction: training in semantic blocks led to interference during acquisition relative to training in unrelated contexts.

Semantic prediction 2: Training in semantic blocks increases the distinctiveness of representations by strengthening the connections between distinctive semantic features and lexical nodes while weakening the connections between shared semantic features and lexical nodes. This prediction was evaluated by examining performance on the semantic probe tasks at two time scales: over the training sessions and at follow-up two weeks after training. According to the predictiction, participants should be slower and/or less accurate to verify shared features of items trained in semantic blocks than to verify distinctive features trained in the same semantic blocks or in other blocking contexts (segmental or unrelated). In both the written and spoken experiments, analyses of the semantic probe task across sessions and at follow-up revealed a relative advantage for distinctive features as compared to shared features. These results provide support for the prediction: semantic blocking during training increased distinctiveness.

Semantic prediction 3: Training in semantic blocks is beneficial for retention relative to training in unrelated contexts, assuming that increased distinctiveness leads to better long-term learning outcomes. This prediction was examined by comparing recall of the names of items trained in semantic blocks to items trained in unrelated contexts. Retention between sessions was considered, as was retention to follow-up two weeks after training. According to the prediction, participants should show better retention of the names of items trained in semantic blocks than of items trained in unrelated contexts. There was some weak evidence to support this prediction: participants as a whole were significantly faster to respond in recall trials at follow-up for the items trained in semantic vs. unrelated blocks in the spoken experiment. However, similar effects were not found in the written experiment or in the analysis of retention over sessions. The significant result suggests that training in semantic blocks may be beneficial for long-term retention as compared to training in unrelated blocks, providing partial support for the prediction.

Semantic prediction 4: Semantic blocking is a desirable difficulty: increased training difficulty (interference) leads to increased long-term retention due to increased distinctiveness. This prediction was assessed in two ways. First, the pattern of results across participants was considered. According to the prediction, participants as a whole should show interference during training but an advantage in retention for items trained in semantic blocks as opposed to unrelated contexts. As described above, there was less improvement for items trained in semantic blocks vs. unrelated contexts over training, and there was partial support in the spoken experiment for better retention of items trained in semantic blocks vs. unrelated contexts at follow-up. This weakly suggests that semantic blocking may be a desirable difficulty. Second, correlations between individual differences in semantic blocking effects on training and retention were considered. According to the prediction, there should be a significant negative correlation between training and retention effects. Results of the written experiment were consistent with this: the participants who demonstrated the most interference during training due to training in semantic blocks vs. unrelated contexts also showed the largest advantage in retention for the names of those items over sessions. Together, these analyses offer some support for the prediction: semantic blocking fits the profile of a desirable difficulty.

Discussion of semantic blocking. The first two predictions regarding semantic blocking were upheld. These are the predictions that directly result from extending the incremental learning model of word production to word learning. As predicted, difficulty during training increased when the items being learned were presented in semantically related as opposed to unrelated blocks. Training in semantic blocks also enhanced the distinctiveness of representations, strengthening distinctive features while weakening shared features, again in accordance with the prediction.

The second two predictions, which link the effects of training to effects on long-term learning, were less strongly supported. Name recall was positively impacted by the increased difficulty of training in semantic blocks, likely as a result of the increased distinctiveness that training conferred on the representations. However, these retention results were relatively weak, applying only to measures of response time at follow-up in the spoken experiment. No recall advantage was found in the written experiment at follow-up or in the analysis of retention over sessions for either the written or the spoken experiment. The weakness of these recall results in the face of the earlier results confirming the first and second predictions casts doubt on the assumption that increased distinctiveness leads to better long-term learning outcomes, although the null results may be a result of lack of power to detect effects not a failure of the model. Similarly, the fourth prediction regarding semantic blocking's status as a desirable difficulty was weakly supported. The overall pattern of training and retention at follow-up in the spoken experiment was consistent with semantic blocking as a desirable difficulty, as was the correlation between individual effects of training and retention across sessions in the written experiment. However, other analyses (e.g., the overall pattern in the written experiement and the correlation in the spoken experiment) were not consistent with this finding, although again these inconclusive results may be due to low power.

Overall, results were strongly consistent with the predictions of the e-ILM that directly extended the model to training, but the predictions that connected training to long-term learning were not unequivocally supported. Although there was some support for the assumptions regarding long-term learning, further investigation is required. In future studies ,retention should be tracked more carefully over time, including assessment at later timepoints, in order to better characterize how training in semantic blocks as opposed to unrelated contexts affects retention. Statistical power should be increased by including a larger number of items and more participants.

Segmental Blocking

The e-ILM predictions also address the effects of training in segmentally related blocks.

Segmental prediction 1: Training in segmental blocks leads to interference during acquisition relative to training in unrelated contexts. As in the evaluation of semantic prediction 1, this prediction was evaluated by examining the interaction between block type and training attempt within session. Here, block type contrasted segmental and unrelated training contexts. According to the prediction, participants should show reduced improvement in speed and/or accuracy during training for items in segmental blocks relative to items in unrelated blocks. Results showed that again, participants did improve with practice for items trained in both contexts, but interference was observed for items trained in related blocks. That is, the improvement within sessions for items trained in segmental blocks was significantly slower than for items trained in unrelated contexts. Significant effects were found in analyses of response time for both the written and spoken experiments. The prediction held for segmental blocking: training in segmental blocks led to interference during acquisition relative to training in unrelated contexts.

Segmental prediction 2: Training in segmental blocks reduces the distinctiveness of representations by strengthening the connections between lexical nodes and shared segments while weakening the connections between lexical nodes and distinctive segments. This prediction was assessed by examining performance on the segment probe tasks over the training sessions and at follow-up. According to the prediction, participants should be faster and/or more accurate to verify shared segments of items trained in segmental blocks than to verify distinctive segments. Although all results did not achieve statistical significance at the $p<.05$ level, there was a consistent pattern whereby participants showed an advantage for shared segments of items trained in segmental blocks relative to distinctive segments. This advantage was present in both the written and spoken analyses over training sessions, but only in written accuracy analyses at follow-up. Overall, however, the general pattern of results observed in this experiment does support the prediction that distinctiveness is reduced when items are trained in segmental blocks.

Segmental prediction 3: Training in segmental blocks is detrimental for retention relative to training in unrelated contexts, assuming that reduced distinctiveness leads to worse long-term learning outcomes. This prediction was examined by comparing recall of the names of items trained in segmental blocks to those trained in unrelated contexts, considering retention both between sessions and at follow-up. According to the prediction, participants should show worse retention of items trained in segmental blocks than of items trained in unrelated contexts. However, results did not support this prediction: there were no significant effects of training in segmental blocks vs. unrelated contexts on retention. There was no retention advantage or disadvantage for training in segmental blocks: results did not conclusively support differences in performance for items trained in segmental and unrelated blocks. The prediction was not upheld; possible explanations for this are discussed below.

Segmental prediction 4: Segmental blocking is not a desirable difficulty: increased training difficulty (interference) leads to reduced long-term retention due to reduced distinctiveness. This final prediction was investigated by looking at the pattern of results across participants and by examining the correlations between individual differences. According to the prediction, participants as a whole may show interference during training, but not an advantage in retention for items trained in segmental blocks as opposed to unrelated contexts. Instead, they should show a retention disadvantage for items trained in segmental blocks as opposed to unrelated contexts. As described above, there was indeed less improvement during training for items trained in segmental blocks vs. unrelated contexts. However, null effects were observed for retention. While segmental blocking did not conclusively fit the pattern of a desirable difficulty, it did not fit the e-ILM prediction either. The prediction regarding correlations between the segmental blocking effects on training and retention state that there should not be a significant negative correlation between training and retention effects as was expected for semantic blocking, but rather a positive correlation. Again, results were only partially consistent with this prediction: there was not a significant correlation between the effects of training in segmental blocks vs. unrelated contexts on training and retention. That is, increased interference during training due to block type was not correlated with retention of those items' names. This is potentially consistent with the idea that segmental blocking is not a desirable difficulty, but it is not consistent with the more specific e-ILM prediction that segmental blocking should increase training difficulty and reduce retention relative to training in unrelated contexts. Null results must be interpreted carefully. Although segmental blocking is not a desirable difficulty, this is because training in segmental blocks as opposed to an unrelated context did not appear to impact retention, not because this type of training had a negative impact on long-term learning outcomes as predicted.

Discussion of segmental blocking. The e-ILM predictions regarding segmental blocking were not consistently confirmed. As for semantic blocking, results were generally consistent with the first two predictions regarding extensions of the incremental learning model from production to training. As predicted, segmental blocking did increase training difficulty relative to training in unrelated blocks. Also as predicted, segmental blocking decreased distinctiveness, resulting in advantages for shared segments vs. distinctive segments on segment probe tasks. However, some of these distinctiveness effects achieved only marginal significance ($p<.1$ not $p<.05$). One reason significant effects may be harder to detect in the segment probe task than in the semantic probe task is that segments are drawn from a closed class. It is necessarily true that the distinctive segments are repeated within the experiment in different blocking contexts. This may reduce differences between them and the shared segments. They are not as distinctive as the distinctive features in the semantic probe experiment, which were not repeated within the experiment. The distinctive features in the semantic probe task truly distinguished between items since they belonged only to one item, while the distinctive segments in the segment probe task were only distinctive within each block since they appeared in other items trained in other contexts within the experiment. Therefore it may be easier to detect differences in shared and distinctive features in the semantic probe tasks than the segment probe tasks.

Results were not consistent with the e-ILM predictions that addressed the relationship between training effects and long-term learning outcomes. Contrary to the third prediction, there was not a negative impact of training in segmental blocks on long-term recall. Why might this be? One possibility is that one of the underlying assumptions of the prediction is wrong. In relating theories of long-term learning and distinctiveness to the production model, I made the assumption that increased distinctiveness leads to better long-term outcomes. This is not necessarily the case. It could be that since participants were able to distinguish between the names of the items they were learning and received indirect feedback in the study portion of the training trials, it did not matter that segmental blocking reduced distinctiveness. The increased difficulty caused by reduced distinctiveness may not lead to reduced recall because of some other mechanism not considered here, such as cognitive control. People may be less susceptible to interference due to segmental overlap than predicted because they must constantly cope with overlapping segments throughout the words of the language, not just in the experimental context. If they could not effectively learn words that share segments, they would have extremely limited vocabularies. Therefore, language users may adapt to overcome interference due to segmental overlap.

Another reason for the failure to find a negative effect of training in segmental blocks on retention might be that distinctiveness was not changed to the extent that was expected in the experiment. As noted above, the pattern of distinctiveness effects showed reductions of distinctiveness, but some had reduced statistical significance. This may indicate that the manipulation of segmental context was less effective than was desired, possibly as a result of items throughout the experiment using many of the same segments. This was necessary because there is a closed set of legal segments in English. Anecdotally, when participants were questioned about their own observations during the experiment at the end of the follow-up session, almost all had noticed that items appeared in blocks that shared meaning, but very few noticed that items appeared in blocks that shared letters or sounds, possibly because so many segments were shared throughout the experiment. While conscious awareness of similarity should not be required to observe effects of it, this does indicate that the manipulation of segmental similarity was less salient than that of semantic similarity and might have had weaker effects. Beyond the sharing of segments across blocks, the manipulation of segmental similarity could be flawed in some way. For instance, perhaps the overlap of phonological features, the consonant-vowel structure of the items, visual similarity, or other potentially relevant dimensions should be taken into account.

A final possible explanation is that the methodology of this study may not have been ideal for observing retention effects. First, statistical power may have been too low to reveal potentially small effects of blocking on retention. More items and more participants may be needed to clearly see such effects. Second, this study may not have tracked retention over a long enough time period. Perhaps the effects on retention take longer to become apparent. For instance, it is possible that items trained in segmental blocks would show an advantage whereby they are retained for longer time periods than those trained in unrelated contexts. Such methodological issues may also explain the weakness of the retention effects found for semantic blocking.

Further research is necessary to distinguish between these alternatives and to more fully characterize the effects of segmental blocking on retention. Future studies should increase power by including more participants and/or training items. They should also more carefully track retention over time, including following participants to later time points where effects might be apparent. Other definitions of segmental similarity should be considered in the design of stimuli, and other methods for reducing distinctiveness should be explored to see if reduced distinctiveness does in fact lead to reduced recall.

Turning to the final prediction, segmental blocking did not fit the profile of a desirable difficulty. Although there was interference due to training in segmental blocks vs. unrelated contexts, there was not a corresponding increase in retention. This result did not support the e-ILM prediction that the difficulty induced by training in segmental vs. unrelated contexts should lead to reduced retention. While the null results are not conclusive, they are potentially consistent with the idea that segmental blocking may not be a desirable difficulty. Not every manipulation that increases difficulty during learning leads to positive long-term learning outcomes; not all difficulties are desirable.

Overall, the results regarding segmental blocking were generally consistent with the e-ILM in terms of effects on training. However, the predictions connecting these effects to long-term learning were not strongly supported. Training in segmental blocks vs. unrelated contexts did not negatively impact long-term learning outcomes even though there was interference in training and reduced distinctiveness. Further research is needed to investigate the cause of the null effect, which may be incorrect assumptions, language-external mechanisms for overcoming interference, or imperfect methodology.

Summary of Predictions and Evidence from This Study

The e-ILM predictions and the results of the study of new word learning in neurotypical adults that bear on these predictions are listed in Table 7.

Table 7. Predictions of the e-ILM regarding the effects of training in semantic and segmental blocks, along with results of Study 1 that were used to evaluate these predictions.

Semantic Blocking

Prediction	Result	
1. Training in semantic blocks leads to interference during acquisition relative to training in unrelated contexts.	Supported: greater improvement within session for items trained in unrelated contexts than in semantic blocks	✔
2. Training in semantic blocks increases the distinctiveness of representations by strengthening the connections between distinctive semantic features and lexical nodes while weakening the connections between shared semantic features and lexical nodes.	Supported: advantage for verification of distinctive features vs. shared features of items trained in semantic blocks on semantic probe tasks	✔
3. Training in semantic blocks is beneficial for retention relative to training in unrelated contexts, assuming that increased distinctiveness leads to better long-term learning outcomes.	Partially Supported: faster responses to items trained in semantic blocks vs. unrelated contexts at follow-up in the spoken experiment	~✔
4. Semantic blocking is a desirable difficulty: increased training difficulty (interference) leads to increased long-term retention due to increased distinctiveness.	Partially Supported: pattern of results across subjects of interference during training and potential retention advantage at follow-up for items trained in semantic blocks vs. unrelated contexts in spoken experiment; correlation of individual differences shows relationship between training difficulty and retention advantage over sessions for items trained in semantic blocks vs. unrelated contexts in written experiment	~✔

Segmental Blocking

Prediction	Result	
1. Training in segmental blocks leads to interference during acquisition relative to training in an unrelated context.	Supported: greater improvement within session for items trained in unrelated contexts than in segmental blocks	✔
2. Training in segmental blocks reduces the distinctiveness of representations by strengthening the connections between lexical nodes and shared segments while weakening the connections between lexical nodes and distinctive segments.	Supported: disadvantage for verification of distinctive features vs. shared features of items trained in segmental blocks on segmental probe tasks	✔
3. Training in segmental blocks is detrimental for retention relative to training in unrelated contexts, assuming that reduced distinctiveness leads to worse long-term learning outcomes.	Not Supported: no significant differences in retention after training in segmental blocks vs. unrelated contexts	✖
4. Segmental blocking is not a desirable difficulty: increased training difficulty (interference) leads to reduced long-term retention due to reduced distinctiveness.	Not Supported: pattern of results across subjects of interference during training but no effect on retention of training in segmental blocks vs. unrelated context; no significant correlation between training and retention effects when individual differences considered.	✖

Overall, the outcomes of the analyses presented in this chapter suggest that the predictions of the e-ILM that directly address training are upheld when the model is extended from production of known words to learning of new words in neurotypical adults. For both semantic and segmental blocking, the predictions about interference during training and changes in distinctiveness were upheld. However, the predictions that relate the model to theories of long-term learning were not consistently supported. There was only weak evidence that training in semantic vs. unrelated contexts led to improved retention relative to training in unrelated blocks, and no evidence of any effect of training in segmental vs. unrelated contexts on retention. In line with the predictions at a broad level, there was evidence that semantic blocking is a desirable difficulty but segmental blocking may not be. However, the contrast between the two was less clear than expected, in part because neither type of blocking affected retention as expected. Support for semantic blocking as a desirable difficulty was equivocal, and segmental blocking did not have the opposite pattern whereby increased difficulty during training in segmental vs. unrelated contexts was expected to lead to reduced retention. Taken together, results suggest that the e-ILM adequately describes the effects of blocking during training, but that these training effects do not impact long-term learning as expected. Further research is needed to better characterize the effects of blocking on long-term learning outcomes.

In the next chapter, I describe another study that investigated whether similar effects are seen in individuals with dysgraphia as they relearn the spellings of words. I evaluate the same testable predictions regarding learning in the context of semantic blocking, segmental blocking, and unrelated items in a clinically relevant population.

CHAPTER 5: STUDY 2. TREATMENT OF SPELLING IN INDIVIDUALS WITH DYSGRAPHIA

Introduction

In the preceding study (Chapter 4), I presented an investigation of the effects of similarity during training of new words in neurotypical adults, testing hypotheses derived from models of the normal production system that address how the different types of similarity affect the trajectories and outcomes of learning as well as the distinctiveness of representations. This chapter presents an extension of this work exploring the effects of similarity during training that has a practical application: helping individuals with dysgraphia relearn the spellings of previously known words. In this study, individuals who exhibit various cognitive deficits affecting spelling were trained on the spellings of words in semantically related, segmentally (orthographically) related, or unrelated blocks. The goal of the study was to examine the effects of blocking on relearning of spellings. Do the effects uncovered in the previous study with neurotypical participants hold for this population of individuals with complex cognitive deficits?

To date, there is limited information about the effects of blocking on the outcomes of treatment for dysgraphia. While there have been a number of investigations of how best to treat spelling deficits, there has been little attention focused on which words should be trained together (though see Sage & Ellis, 2006 for one study exploring treatment of orthographic neighbors, summarized in Chapter 2). There is also some limited evidence that training words for spoken production in semantically related sets can be beneficial for individuals with aphasia who do not have deficits affecting the semantic system. This has been explained as strengthening the lexical-semantic representations, which presumably allows stronger activation to cascade through the segmental encoding process (Laine & Martin, 1996; Martin & Laine, 2000). However, there has not been direct comparison of treating semantically related, segmentally related, and unrelated words, and these studies did not address spelling but rather spoken production.

What sort of effects might blocking by semantic or segmental similarity have for individuals with cognitive deficits impacting spelling processes? One possibility is that the same effects seen for neurotypical adults learning new words may hold since these individuals with dysgraphia are attempting to incorporate word knowledge into production systems that were premorbidly similar to those of the neurotypical adults. Another possibility is that the level of deficit interacts with the effect of blocking, so that participants with deficits affecting lexical selection have different effects for semantic and/or segmental blocking relative to participants with deficits affecting segmental encoding. Deficits may shift the delicate balance between interference and facilitation that leads to the effects seen in neurotypical participants, so different directions of effects are possible. A different possibility concerns the distinctiveness of representations. If damage impacts representations in such a way that similar items are no longer distinctive and it is difficult for participants to make distinctions, training similar items together may exacerbate the problem and lead to interference both during acquisition and when recall is measured at a later time. Alternatively, training in blocks may help participants make distinctions through direct comparison of similar items, which would instead lead to positive outcomes. A related possibility is that blocking may allow for strategy use (e.g., realizing that all words in a segmental block contain a particular letter) that may facilitate training but not necessarily retention.

Thus there are a variety of possible outcomes that may result from training individuals with dysgraphia on words that are related in meaning or form. The group of individuals who participated in the present study had a range of complex deficits of spelling. While these individuals are typical of the clinical population in need of treatment, they are not ideal for investigating specific relationships between specific deficits and treatment outcomes. However, this study does provide an opportunity to see whether the predicted effects of the e-ILM proposed in Chapter 3 and tested in neurotypical learners in Chapter 4 apply to this population as well. In addition to providing an additional test of the theoretical predictions that have implications for understanding the production system, the results of the present study with individuals with dysgraphia may have clinical implications for the context in which spelling should be treated.

In the present study, seven individuals with dysgraphia were trained on matched sets of words that were semantically related, segmentally related, or unrelated. Performance during training was analyzed to investigate the predictions that both semantic and segmental blocking lead to interference during acquisition relative to training in unrelated contexts. Performance immediately after training as well as at follow-up several weeks later was analyzed to examine the impact of the training context on retention (long-term learning), investigating the predictions that semantic blocking leads to improved retention relative to training in unrelated contexts and that segmental blocking leads to reduced retention relative to training in unrelated contexts. The pattern of results for the group across training and retention as well as analyses of individual differences allow for evaluation of the predictions that semantic blocking is a desirable difficulty while segmental blocking is not. The study did not probe the distinctiveness of representations, so the predictions that semantic blocking increases distinctiveness while segmental blocking reduces it were not directly addressed. Overall, this study asked whether the e-ILM predictions accurately describe the effects of training in semantic and segmental blocks as opposed to unrelated contexts for this group of individuals with dysgraphia.

Case Histories

Participants

Spelling is often neglected in speech therapy in favor of immediate communicative needs for spoken language, and speech language pathologists typically have little formal training in therapies for it. However, as individuals with aphasia recover, they often express a desire to work on spelling, as written language comprises an important part of modern life through such forms of communication as texting, email, and social media participation, as well as providing an additionally modality to augment spoken communication. A recent survey of stroke survivors shows that spelling is listed as one of the top five residual problems by 71% of those who had left hemisphere strokes and by 71% of their caregivers (Hillis & Tippett, 2014).

In the present study, seven participants who expressed interest in practicing spelling were recruited from a local community based aphasia support center. All had experienced strokes that resulted in language deficits at least three years prior to the study. Demographic information about these seven individuals is presented in Table 8.

Table 8. Demographic information about the participants.

	KSR2	*GHN*	*SMY*	*REN*	*ESG*	*DWS*	*DDR*
Age	50	68	56	54	60	63	72
Handedness	right	right	right	left	right	right	right
Profession	computer programmer	physician	architect	medical library administrator	mechanic/ driver	newspaper printer	engineer
Years post-stroke	5	4	2 strokes: 8 & 4 years prior	20	3	9	3

Background Spelling Assessment

The first step in undertaking this treatment study was to obtain some background information about the spelling deficits of the seven participants. Although this particular group was not ideal for investigating the relationships between specific deficits and the effects of treatment because multiple components of the spelling system are damaged for these individuals, it is still important to know what sort of spelling problems these individuals face. This may be helpful in the present study for interpreting the effects seen here and in future studies that may extend the work to other individuals with potentially similar deficits.

In characterizing spelling deficits, it is not common to discuss lexical selection and segmental encoding processes, the processes that have been the focus of earlier discussion in this dissertation. Instead, spelling deficits are typically described as impairments of specific components of the spelling process. However, as discussed in Chapter 1, lexical selection and segmental encoding can be thought of as subcomponents of orthographic and phonological long-term memory (i.e., they are processes that take place within the component of the lexicon). In the following discussion, I will briefly review the cognitive architecture of written production that was initially presented in Chapter 1 and discuss the consequences of damage to the various components for behavioral performance. I will then provide information regarding the participants' cognitive deficits.

Cognitive architecture of spelling and its disorders. A cognitive architecture of spelling is presented in Figure 28. This model and the following discussion are in line with cognitive architectures previously described in the literature (for in-depth reviews of the spelling system and consequences of damage to the various components, see Buchwald & Rapp, 2009; Ellis & Young, 1988; Rapp & Gotsch, 2001;Rapp, 2002; Tainturier & Rapp, 2001).

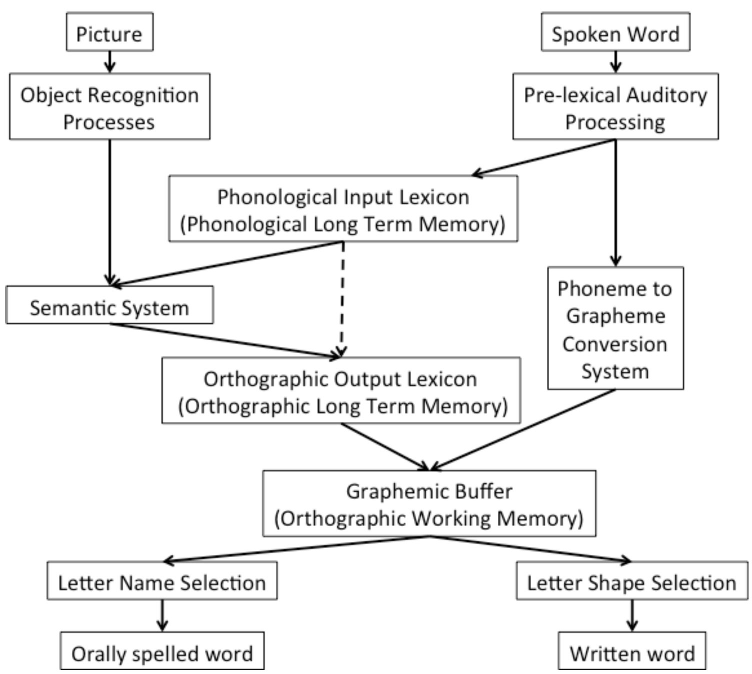


Figure 28. A cognitive architecture of spelling

I begin by describing the processes involved in spelling a word to dictation and the behavioral consequences of damage to the components of the system. First, prelexical auditory processing converts the auditory stimulus of the spoken word to a phonological representation. If there is damage to the prelexical auditory processing, spelling to dictation will be impaired.

From this point, spelling can happen via two paths: the sublexical route and the lexical route. When the sublexical route is used, the stimulus's phonemes are directly converted to graphemes, making use of stored, regular mappings of phonology to orthography. If the phoneme-to-grapheme conversion system is damaged, nonword spelling will be impaired. If an individual hears a stimulus that does not correspond to a known word, he or she cannot rely on the stored representations of the lexical route (discussed below) and will try to sound it out by using phoneme-to-grapheme conversion in the sublexical route. If this route is damaged, this will not be possible: the individual will have marked difficulty in spelling nonwords since neither route can process them.

Alternatively, the lexical route can be used to spell words to dictation. After prelexical auditory processing, the phonemes of the stimulus activate the long-term phonological memory representation in the phonological input lexicon. If the phonological input lexicon is damaged, the individual will have difficulty spelling real words to dictation and understanding spoken words since the stored spoken form of the word is inaccessible. The individual may also show frequency effects whereby responses to high frequency words are more accurate than responses to low frequency words since the representations of words that are encountered more often are more likely to be more resilient to damage. However, if this is the only component that is damaged, the individual will be able to access information about the spelling and meaning of real words via other modalities. For instance, written picture naming will be intact.

Next, the meaning of the word is activated in the semantic system. If the semantic system is damaged, the individual will have difficulty with all tasks that require access to word meanings, regardless of the modality. Spoken and written production and comprehension of real words will be impacted. In spelling, semantically related errors may be produced (i.e., words that are related in meaning to the target; e.g., cat→dog).

Once the meaning is accessed, activation is sent to the orthographic output lexicon, which stores the orthographic long-term memory representation of the word's spelling. (Note that a direct route is also hypothesized whereby activation can be sent directly from phonological long-term memory to orthographic long-term memory, bypassing the semantic system.) According to the assumptions made in this dissertation, the processes of lexical selection (choosing the correct orthographic long-term memory representation) and segmental encoding (retrieving the graphemes that make up the orthographic long-term memory representation) occur within this component. If orthographic long-term memory is damaged, frequency effects are likely to be observed. Words that are encountered more often are likely to have stronger representations that are more resilient to damage, and so individuals with damage to this component are more likely to spell words correctly if they have higher frequency in the language. Damage to orthographic long-term memory can also lead to semantic errors: if the representation of the target item is damaged, shared features may lead to related lexical nodes having higher activation and being selected. Additionally, damage to this component may result in letter errors (substitutions [e.g., cat→cap], additions [e.g., cat→cfat], and/or deletions of letters [e.g., cat→ct] as well as transpositions and movements of letters [e.g. cat→cta or tca]): damage may result in imprecise or incomplete representations of the form of the word, making selection of the identity and/or serial position of the constituent graphemes inaccurate.

Damage to any component of the lexical route may lead to production of phonologically plausible spelling errors in spelling to dictation if the individual compensates for the damage by relying on the relatively more intact sublexical route. That is, if an individual cannot access the stored representation of a word (either meaning or form), they may convert the phonemes of the word directly to graphemes, resulting in an error made up of graphemes with high probability given the phonemes of the word (e.g., yacht→yot). For the same reason, damage to the lexical route may lead to a phoneme-grapheme probability effect, with words that have more consistent phoneme-to-grapheme mappings spelled correctly via the sublexical route while those with less probable mappings are spelled incorrectly since the most plausible or frequent mapping leads to an error.

The sublexical and lexical routes converge on the orthographic working memory. The output of both processes is a string of abstract graphemes that is held in orthographic working memory. The store of orthographic working memory is often referred to as the graphemic buffer, and it maintains activation for the abstract representations of graphemes being spelled throughout production of that word, making the identity and serial order of individual graphemes available to downstream processes (Buchwald & Rapp, 2009). Damage to the graphemic buffer leads to word length effects: there is higher accuracy for letters in short words than for letters in long words because short words have fewer graphemes that must be maintained in the resource-limited working memory system. Because the representation at this level of processing includes information about the serial position and identity of single graphemes, damage results in order errors involving transpositions and movements of graphemes as well as individual letter errors including substitutions, deletions, or additions of graphemes (as discussed above). Geminate errors, including doubling of the wrong letter (rabbit→rabbitt or rabitt), producing only a single letter instead of a double letter (e.g. rabbit→rabit), or substituting the wrong letter identity for both letters in a double (e.g. rabbit→rannit), also arise after damage to this component.

The orthographic working memory representation can be used to write or say the letters that make up the word. When the word is to be written, letter shape selection generates shapes for the serially ordered letters that are to be written, and motor programs are activated that allow the letter shapes to be written. When the letters of the word are to be spoken aloud, letter name selection generates the spoken names for the serially ordered letters, and motor programs are activated that allow the letter names to be pronounced. Damage to letter shape selection (and/or subsequent motor programs) leads to selective difficulty with written spelling in comparison to oral spelling, whereas damage to letter name selection (and/or subsequent motor programs) leads to the opposite pattern of difficulty.

The processes involved in written picture naming are very similar to those involved in the lexical route of spelling to dictation. Instead of the prelexical auditory processing that happens at the beginning of the spelling to dictation task, object recognition processing occurs when a picture is viewed in order to analyze the visual stimulus. If visual object recognition processes are damaged, written picture naming will be impaired. Once the picture is recognized, it activates the stored representation of its meaning in the semantic system. From this point forward, written picture naming utilizes the same components as the lexical route of spelling to dictation. Deficits affecting the semantic system, orthographic output lexicon, graphemic buffer, letter shape selection and letter name selection have the same consequences for written picture naming as they do for spelling to dictation. Damage to any of these components will cause difficulty in both tasks. Differential performance on the two tasks suggests that some component that is not utilized in both is damaged. If prelexical auditory processes are damaged but visual object recognition processes are intact, there will be better performance for written picture naming than spelling to dictation and vice versa. If only the phonological input lexicon is damaged, performance on spelling to dictation will be worse than performance on written picture naming, especially for words that have low probability phoneme-to-grapheme mappings that would lead to errors when the sublexical route is used. If the phoneme-to-grapheme conversion system (sublexical route) is damaged, spelling to dictation may be impacted, but this may not be visible unless there is also damage to the lexical route or if nonwords are used as stimuli so that the lexical route cannot compensate.

Damage to each of the components of the spelling system leads to observable effects as described above. In the following section, I will present information about the performance of the individual participants from this study on a battery of screening assessments that may provide information about their deficits. Damage to the components of the lexical route, especially the processes of lexical selection and segmental encoding, may be related to the effects of semantic and segmental blocking for individuals with deficits. For example, predictions about semantic blocking concern input from the semantic system and processing within orthographic long-term memory, while predictions about segmental blocking concern processing within orthographic long-term memory and output to orthographic working memory. Individuals with different damage to these components may show differences in the effects of blocking as compared to neurotypical individuals. Therefore, information about the participants' deficits may be helpful in interpreting the relationship between data observed in the treatment study and the e-ILM predictions.

Pre-treatment assessments of spelling. In order to characterize pre-treatment spelling performance, a battery of neuropsychological assessments was administered. Table 9 presents performance on these tasks. The accuracy measure used in this chart is letters correct, not whole responses correct. Many of the participants made few correct responses, so looking at whole responses might mask subtle effects that are visible when examining letters correct. Classification of errors is presented in Table 10. Below, the battery of assessments and the patterns of performance of the participants are discussed.

Table 9. Performance on screening assessments. Accuracy is reported as letters correct since few whole responses were produced correctly for many of the participants. Significant differences (α=0.05) in the types of words being compared are bolded.

Modality	Assessment	Effect		DDR	DWS	ESG	GHN	KSR2	REN	SMY
spelling to dictation	Length List	Frequency	high frequency	4/84 (5%)	15.5/84 (18%)	71.25/140 (51%)	32.5/84 (39%)	28/140 (20%)	66.75/140 (48%)	46/53 (88%)
			low frequency	5/84 (6%)	5/84 (6%)	39.35/140 (28%)	19/84 (23%)	25/140 (18%)	42/140 (30%)	24/44 (55%)
			analysis	X^2=0.12, p=0.732	**X^2=6.13, p=0.013**	**X^2=15.31, p<0.001**	**X^2=5.10, p=0.024**	X^2=0.21, p=0.647	**X^2=9.208, p=0.002**	**X^2=13.88, p<0.001**
		Length	short words (3&4 letters)	4/56 (7%)	8/56 (14%)	52.25/98 (53%)	21/56 (38%)	16.25/98 (17%)	50/98 (51%)	25.5/32 (80%)
			long words (5&8 letters)	5/112 (4%)	12.5/112 (11%)	58.25/182 (32%)	30.5/112 (27%)	36.75/182 (20%)	58.75/182 (32%)	44.5/64 (70%)
			analysis	X^2=0.47, p=0.528	X^2=0.34, p=0.560	**X^2=12.11, p<0.001**	X^2=1.85, p=0.173	X^2=0.54, p=0.462	**X^2=9.42, p=0.002**	X^2=1.11, p=0.291
written picture naming	PALPA39	Length	short words (3&4 letters)	did not attempt	30.75/42 (73%)	32/42 (76%)	40/42 (95%)	11.25/42 (27%)	34/42 (81%)	did not attempt
			long words (5&6 letters)		21/66 (32%)	25.5/66 (39%)	43.25/66 (66%)	15/66 (23%)	51.5/66 (78%)	
			analysis		**X^2=17.62, p<0.001**	**X^2=14.54, p<0.001**	**X^2=12.82, p<0.001**	X^2=0.23, p=0.632	X^2=0.13, p=0.715	
	PALPA54	Regularity	regular words	10.5/52 (20%)	42.5/100 (43%)	40.75/100 (41%)	55.25/100 (55%)	28.5/100 (29%)	62.5/100 (63%)	46/52 (88%)
			irregular words	5/45 (11%)	37.4/96 (39%)	33.25/96 (35%)	58.25/96 (61%)	22.5/96 (23%)	68/96 (71%)	33.75/45 (75%)
			analysis	X^2=1.48, p=0.224	X^2=0.25, p=0.614	X^2=0.78, p=0.377	X^2=0.59, p=0.442	X^2=0.65, p=0.419	X^2=1.53, p=0.216	X^2=2.99, p=0.084
picture matching	Pyramids & Palm Trees	Object semantics	score	did not attempt	14/14 (100%)	11/14 (79%)	14/14 (100%)	13/14 (93%)	14/14 (100%)	did not attempt

Table 10. Classification of responses from screening assessments into error types. Columns list the number of responses that fit each category, with the percentage of all trials that fit that category in parentheses.

	Participant	DDR	DWS	ESG	GHN	KSR2	REN	SMY
	total responses	48	92	120	92	120	120	36
	correct responses	0 (0.0%)	10 (10.9%)	16 (13.3%)	22 (23.9%)	0 (0.0%)	42 (35.0%)	15 (41.7%)
	omission errors	7 (14.6%)	19 (20.7%)	2 (1.7%)	1 (1.1%)	0 (0.0%)	2 (1.7%)	0 (0.0%)
whole word errors	semantic word errors	0 (0.0%)	15 (16.3%)	4 (3.3%)	2 (2.2%)	13 (10.8%)	21 (17.5%)	0 (0.0%)
	morphological word errors	0 (0.0%)	1 (1.1%)	0 (0.0%)	1 (1.1%)	0 (0.0%)	2 (1.7%)	1 (2.8%)
	other word errors	4 (8.3%)	2 (2.2%)	11 (9.2%)	10 (10.9%)	16 (13.3%)	5 (4.2%)	2 (5.6%)
	total word errors	**4 (8.3%)**	**18 (19.6%)**	**15 (12.5%)**	**13 (14.1%)**	**29 (24.2%)**	**28 (23.3%)**	**3 (8.3%)**
nonword errors	likely semantic nonword errors	1 (2.1%)	5 (5.4%)	6 (5.0%)	8 (8.7%)	10 (8.3%)	13 (10.8%)	1 (2.8%)
	phonologically plausible nonword errors	0 (0.0%)	0 (0.0%)	2 (1.7%)	1 (1.1%)	1 (0.8%)	2 (1.7%)	1 (2.8%)
	other nonword errors	36 (75%)	40 (43.5%)	79 (65.8%)	47 (51.1%)	78 (65%)	33 (27.5%)	16 (44.4%)
	total nonword errors	**37 (77.1%)**	**45 (48.9%)**	**87 (72.5%)**	**56 (60.9%)**	**89 (74.2%)**	**48 (40.0%)**	**18 (50.0%)**

Description of the screening battery. This screening battery consisted of both spelling to dictation and written picture naming tasks. On spelling to dictation tasks, the experimenter said a word aloud, and the participant was to write the word. In this study, the four and eight letter length list from the JHU Dysgraphia battery (Goodman & Caramazza, 1985) was administered. This assessment consists of twenty-eight bisyllabic words. Half of the items are low frequency, and half are high frequency. Half of the words have four letters, and half have eight letters. Nouns, verbs, and adjectives are included among the stimuli. Three of the participants also completed the three and five letter length list from the same JHU Dysgraphia battery, which is also a twenty-eight item list contrasting high and low frequency words. On this list, half of the items have three letters and half have five letters. For these three participants, results for the two lists were combined. Words with 3 or 4 letters were considered short words, and words with 5 or 8 eight letters were considered long words. Data from these lists are be used to investigate frequency effects, which indicate damage to long-term memory (either the phonological input lexicon or orthographic output lexicon), as well as length effects, which indicate damage to orthographic working memory.

The screening battery also included two written picture naming tasks, in which participants were presented with a line drawing of an object and asked to write the name of the picture. The first task was PALPA39 (Kay, Lesser, & Coltheart, 1992), in which participants name 3-6 letter monosyllabic words matched in frequency, imageability, and morphemic complexity. The task was originally designed as a spelling to dictation task, but in this study pictures of each stimulus were selected and presented to participants. Table 9 includes comparison of performance on short 3-4 letter words and long 5-6 letter words. Length effects here are again consistent with damage to orthographic working memory.

The final spelling screening assessment used in this battery was PALPA53 (Kay et al., 1992), which compares spelling of words with regular and irregular grapheme-phoneme mappings[17] that are matched on frequency, familiarity concreteness, age of acquisition, letter length, and number of syllables (Kay et al., 1992). Regularity effects are assumed to indicate lexical deficits. Since regular words can be spelled via the sublexical route but irregular words cannot, showing an advantage for regular words over irregular words suggests that participants are compensating for deficits in the lexical route by using the sublexical route to spell these words.

Participants also completed one cognitive assessment that did not require spelling: the short-form Pyramids and Palm Trees task (Breining et al., 2015). This is a 14-item assessment of nonverbal object semantics. On each trial, participants see three pictures and are asked to match the reference object (e.g., GLASSES) to the one of the other two objects that is more associated with it (e.g., target: EYE vs. distractor: EAR). This assessment is designed to quickly identify semantic impairments. The normal range of performance is correct responses on 12-14 out of 14 trials. Performance below this level is indicative of a semantic deficit. Finding such a deficit in these participants would suggest that deficits in spelling may be caused by damage to the semantic system.

Description of error classification. The types of errors that participants made across the screening assessments may also be informative. Responses across all screening assessments were classified in Table 10. This table shows the number and

[17] Note that regularity in PALPA53 refers to regularity of reading, not spelling: this assessment is designed to compare spoken and written picture naming, reading, and repetition.

percentage of responses that fell into several categories. Correct responses were accurate spellings of the target at the whole word level. Omission errors were trials in which the participant did not write any response.

Whole word errors were response in which the participant produced a correctly spelled English word that was not the target on that trial. Within the set of whole word errors, semantic errors were responses in which a participant produced a word related in meaning to the target (e.g., elbow→knee). Also within the set of whole word errors were morphological errors, which were responses in which participants produced the wrong grammatical form of the target (e.g., iron→ironing). Because it is hard to define what counts as enough similarity to the target to be a phonologically or orthographically related word error, other whole word errors included those that had some phonological and/or orthographic overlap with the target as well as those that had no obvious relationship with the target. Overall, whole word errors are likely to result from deficits affecting the selection of the correct lexical node. If the target lexical node cannot be accessed, either because the semantic representation is somehow damaged and does not lead to the greatest activation of the target lexical node or because the target lexical node itself is damaged, another lexical node that is highly active may be selected. This is likely to be a semantically related item that is active because it shares features with the target or a form related item that is active through feedback from shared segments. Whole word errors are consistent with deficits in orthographic long-term memory affecting the orthographic output lexicon, although they are also consistent with semantic deficits.

Nonword errors were responses in which the letters produced by the participant did not correspond to a correctly spelled English word. These errors included additions, deletions, transpositions, and substitutions of letters, and are consistent with deficits to the graphemic buffer that affect holding graphemes in working memory and/or with damage to the long-term memory representations stored in the orthographic output lexicon that include information about graphemes. Within the set of nonword errors, likely semantic errors were those that closely resembled other English words that were related to the target. That is, they were potentially misspellings of whole semantic errors (e.g., bread → loal [possible loaf]). These errors may reflect multiple deficits affecting both selection of the appropriate lexical item and production of the selected lexical item's graphemes. Phonologically plausible errors were nonword errors that could be produced by applying phoneme-to-grapheme mappings to the target. That is, they are possible alternative spellings of the words that might be produced through use of the sublexical route (e.g., ready→reddy). Overall, phonologically plausible errors point to deficits in the lexical route, reflecting compensatory use of the sublexical route. Few phonologically plausible errors are observed when both the lexical and sublexical routes are damaged. Other nonword errors include those that have some orthographic or phonological overlap with the target as well as those with no obvious relationship to the target.

Discussion of participant performance on screening assessments. Across tasks, participants exhibited a wide range of performance indicating a variety of spelling deficits.

The participants in this study all appeared to spell using a damaged lexical route. None of the participants in this study exhibited a significant regularity effect. Although the integrity of the sublexical route was not directly tested through assessment of nonwords, impairment of this route was suggested by the observation that participants produced very few phonologically plausible errors. Additionally, no participant showed a significant effect of regularity. An advantage for regular words as compared to irregular words would have suggested that patients were compensating for lexical deficits via the sublexical route, since the correct spellings of regular words, but not irregular words, can be produced via the sublexical route. The failure to find regularity effects suggests that participants did not use semantic information to extract the long term phonological memory representation of the words and then use the sublexical spelling route to convert phonemes to graphemes, a possible compensatory mechanism to overcome deficits in the lexical route.

In examining the lexical route, performance on the Pyramids and Palm Trees task showed that these participants did not display gross deficits of the semantic system. Only ESG performed below the normal range (12-14/14) with a score of 11/14, and his difficulty may have been due to difficulty understanding the task as opposed to semantic damage.

The majority of participants exhibited damage to orthographic long-term memory. One clear indication of this was the significant frequency effects displayed by five out of seven participants in spelling to dictation. Frequency effects indicate damage to long-term memory representations. Individuals with damage to either the phonological input lexicon or the orthographic output lexicon are likely to spell high frequency words more accurately than low frequency words. This is because the often-encountered words are likely to have stronger representations that are less susceptible to damage. In this particular group of participants, damage to the orthographic output lexicon (orthographic long-term memory) was indicated by the observation that performance was similar across spelling to dictation and written picture naming tasks[18]. If it were only damage to the phonological input lexicon (phonological long-term memory), better performance would be expected on written picture naming tasks than on spelling to dictation tasks since written picture naming does not utilize the phonological input lexicon. Since performance on the two types of tasks was similar, damage to orthographic long-term memory is indicated (although this does not rule out the possibility that participants could have deficits in phonological long-term memory as well).

The majority of participants also showed damage to orthographic working memory. Four of the seven participants showed significant effects of length on at least one assessment that compared long and short words. A fifth, SMY, showed a trend in the same direction on the spelling to dictation assessment. The failure to find a significant length effect for this participant may have been due to lack of power: he completed only half of the four and eight letter length list and did not attempt PALPA39. The length effects observed for these participants indicate damage affecting the resource-limited orthographic working memory system.

Five of the seven participants thus showed complex deficits of the lexical route that indicated damage to both orthographic long-term memory and orthographic working

[18] Note that participant DWS did show somewhat better performance on written picture naming than on spelling to dictation, but this was because the spelling to dictation task was administered earlier in testing before it was discovered that he was often unwilling to guess on these screeners but did respond to cuing when the experimenter provided the first letter.

memory. What about the remaining two participants? The first of these was DDR. It was very difficult to characterize DDR's deficit because he produced so few letters in his responses: only 1-2 were attempted for the majority of responses, and there were also many omissions in which no response was attempted. The letters that were produced were often not part of the intended target. The limited responses may mask potential underlying effects; it is hard to say with any confidence which components of the lexical route were intact vs. impaired.

The other participant who diverged from the general pattern was KSR2. KSR2 produced sufficient responses, but did not show frequency or length effects. His deficit seemed to be specifically in accessing the correct orthographic long-term representation of the target word. Across tasks, although he was not the only participant to make semantic errors, he did commit many of them (e.g., seven → six) as well as several unrelated whole word substitutions (e.g., arm → then) that are consistent with this sort of difficulty. Further evidence comes from the contrast between his performances on written and spoken naming. While performing the written task, he often correctly said the word aloud while writing something else, indicating that he had difficulty retrieving the correct orthographic long-term representation for the word, even though he could retrieve the correct phonological long-term representation given the semantic information available. Although he made many letter errors leading to nonword responses, the failure to find a length effect suggested that his difficulties arose at the level of orthographic long-term memory. Like the other participants, KSR2 primarily used his damaged lexical route to spell: he made only one phonologically plausible error, indicating that he was not compensating for the damage to the lexical route by spelling with the sublexical route.

The contrast between KSR2 and the other participants suggests that his deficit may predominately affect lexical selection, while the others have deficits that affect segmental encoding and orthographic working memory processes (note that some also have deficits impacting lexical selection).

Overall, the participants in this study showed a number of complex deficits implicating damage to multiple components of the spelling system. Although they are representative of the clinical population that may seek treatment for spelling deficits, they are not an ideal group for investigating the different effects of treatment for those with different deficits. However, data from this group can still be used as a way to test the e-ILM predictions in a new population. It may also be possible to make some preliminary speculations about relationship between deficits and outcomes that can be investigated further in future studies. Below, I describe the treatment study that was used to examine how training in semantically or segmentally related blocks affected treatment outcomes and provided additional data to evaluate the e-ILM predictions.

Spelling Treatment

A timeline of the treatments study is presented in Figure 29 to provide context for the description below.

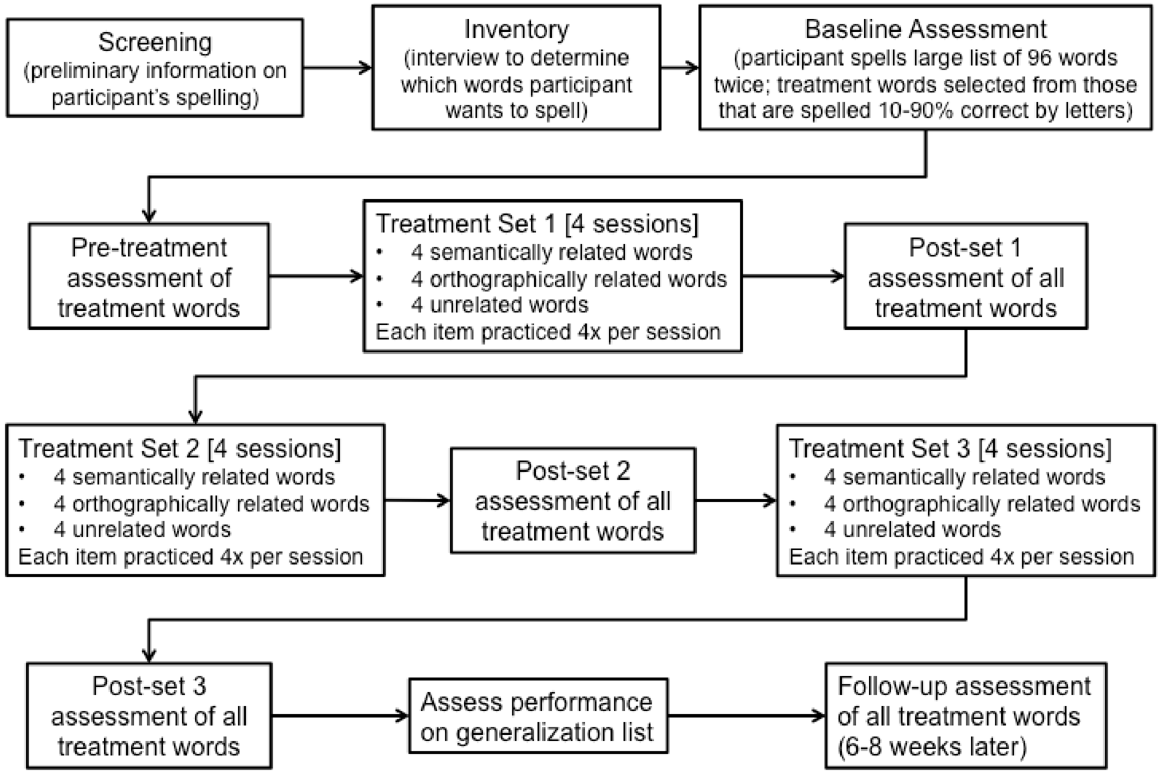


Figure 29. Timeline of treatment study

Selection of Stimuli

After the battery of spelling screeners described above was administered, participants were interviewed about their interests to ensure that treatment lists would include at least some words that were relevant to their needs. To assist in this process, students working at the aphasia treatment center as part of their training to become speech language pathologists developed an inventory that included words from a variety of domains that might be of interest to the participants (e.g., foods, words related to healthcare, sports). The inventory was administered to each participant individually.

Based on their responses to the inventory and performance on the screening battery, a baseline set of stimuli was selected. This consisted of eight words from each of four different semantic categories, eight items from each of four groups that predictably shared orthography, and thirty-two additional unrelated words, for a total of ninety-six words. Groupings were tailored to each participant's interests. For example, REN enjoyed watching sports on television, so one of her semantically related blocks consisted of the names of various sports. GHN expressed a desire to be able to write a grocery list, so one of his semantically related blocks included foods he was likely to purchase. Several participants wanted to be able to write "doctor" due to having many appointments, as well as lists for the drug store where they might purchase items such as a razor, so they had segmentally related blocks where all words ended in –or. Note that semantic and segmental overlap were predictable in this study, not distributed as in the study with neurotypical participants: all items in semantic blocks were from the same category, and all items in segmental blocks shared the same letters. All words were concrete, depictable nouns. A picture was selected for each word on the list.

Over several weekly sessions, the baseline was administered twice to each participant. On each trial, a picture was presented on a single sheet with a line beneath it. The participant was instructed to write the name of the picture on the line. If the participant indicated that he or she was unsure about the name of the picture, the experimenter said it for them. This was allowed because some of the pictures were ambiguous (e.g., "couch" could equally be called "sofa"). Items on the baseline were scored by percentage of letters correct (following Caramazza, Miceli, Villa, & Romani, 1987).

Stimuli were selected based on performance on the two baselines. The goal was to identify 36 words for each participant that were spelled neither completely correctly nor completely incorrectly on both of the two baselines. In order to match treatment blocks on similarity, this constraint was relaxed for some participants. Across all treatment words, mean letter accuracy across the two baseline assessments for each participant ranged from 34.8% to 61.3%.

Each participant's resulting thirty-six treatment words were divided into three sets of twelve words. Within each set, there was a block of four semantically related words, a block of four segmentally related words, and a block of four unrelated words. These three blocks were matched to each other on frequency, length (as measured by number of letters, number of phonemes, and number of syllables), imageability, concreteness, and age of acquisition using measurements from the MRC Psycholinguistic Database (Coltheart, 1981). Selected stimuli for each participant appear in Appendix E.

Treatment

Pre-treatment assessment. After selection of the treatment words, a pre-treatment assessment was administered. Participants attempted to name each of the thirty-six treatment pictures.

Treatment phase. In the treatment phase, the three sets of twelve words were trained separately. First, there were four training sessions for the first set of words (i.e., the same twelve words were trained in four sessions while the other twenty-four treatment words that made up the second and third sets were not trained). This was followed by assessment of all thirty-six treatment items. Next, training of the second set commenced, comprising of four training sessions for Set 2 words. This was again followed by assessment of all treatment words. Finally, the third and final set of words was trained for four sessions, followed by assessment of all thirty-six treatment words.

The treatment consisted of a Test-Study-Test procedure that can be considered a modified version of Copy and Recall Treatment (CART). This method was chosen because it has been shown to be effective at training the representations of specific words in a variety of deficits involving the lexical route of written production (e.g, Ball, de Riesthal, Breeding, & Mendoza, 2011; Beeson & Egnor, 2006; Beeson, Hirsch, & Rewega, 2002; Beeson, Rising, & Volk, 2003; Beeson & Rapcsak, 2012; Beeson, 1999, 2004; Orjada & Beeson, 2005; Rapp & Kane, 2002; Raymer, Cudworth, & Haley, 2003). The procedure was as follows: (1) The participant was presented with a piece of paper, the bottom half of which was covered. At the top of the page, there was a picture depicting one of the treatment items. The participant was instructed to write the name of the picture on a line directly under the picture to the best of his or her ability. (2) The bottom half of the page was uncovered. Here, the name of the picture was printed in upper case letters. The participant was instructed to copy the name of the picture on the line. Use of upper case or lower case was permitted. (3) Steps 1 and 2 were repeated for the same word, using a new sheet of paper.

Within a treatment session for one of the three sets, each block (semantically related, segmentally related, and unrelated) was presented separately. There were two cycles of each block, meaning that all four words were administered as described above and then the same four were administered again in a different order. After two cycles through a block, treatment of the next block began. There were 48 total attempts to name pictures in a session: four attempts for each item (when counting both attempts described in the procedure above separately). Each session lasted a maximum of one hour,[19] and typically took 35-55 minutes to complete. Participants completed 1-2 sessions per week, depending on their schedule of attendance at the aphasia center. While the goal was to complete the same number of sessions each week, some participants had longer breaks between some sessions due to absences caused by illness, travel, and family obligations. Across sessions, the order of blocks varied (e.g., session one might have the semantically related block, followed by the segmentally related block, and then the unrelated block; whereas session two might have the reverse order).

Generalization assessment. Once all thirty-six words had been trained and the immediate post-treatment assessment administered, generalization to untrained items was tested. For each participant, an individualized generalization list was created to compare performance on the same words before and after treatment. Each participant's generalization list was created by combining the list of items from the baseline assessments with items from the screening assessments that were administered prior to treatment, excluding words that were outside the range of frequency and length of the treatment words. For each participant, the generalization list included 171-396 untrained items, 28-60 of which were semantically related to items trained in semantic blocks and 33-46 of which were segmentally related to items trained in segmental blocks. (The range in the number of items depended on the number of items administered in screening assessments from the same range of length and frequency as the trained items.) The resulting generalization list was randomly ordered and administered to participants. This list included both written picture naming trials and spelling to dictation trials. Items that

[19] For some participants, only two of the three blocks were administered in a session due to time constraints. When this happened, additional sessions were added. Each block was trained in four separate sessions.

were written to dictation in the initial screening assessment appeared in spelling to dictation trials, while items that written in response to pictures in the initial screening assessment and in the baseline assessments appeared in written picture naming trials. This was done so that comparisons of performance before and after treatment would rely on information collected from the same task.

Follow-up assessment. Eight to ten weeks (55-70 days) after the immediate post-treatment assessment, a follow-up assessment consisting of all thirty-six treated words was administered.

Results and Discussion

By looking at the patterns of results across these analyses, it is possible to see if the e-ILM predictions (listed in Table 4 in Chapter 3) hold for the individuals with dysgraphia who participated in this study. These predictions suggest that both semantic and segmental blocking increase difficulty during training, but that semantic blocking improves retention while segmental blocking reduces retention. Semantic blocking is predicted to be a desirable difficulty, while segmental blocking is not. The analyses of training and assessment data presented here directly address these predictions. These predictions will be discussed in turn. Note that the e-ILM also predicts that training in semantic blocks enhances distinctiveness, while training in segmental blocks reduces distinctiveness. However, this prediction was not directly addressed in the present study: shared and distinctive features and segments were not probed for comparison.

Data from all seven individuals were entered into the same group mixed effects models.[20] One reason for this is that models of multiple participants have increased power to reveal effects across participants. Another reason was that one of the goals of the analyses presented here is to investigate the relationship between training and retention difficulty. Analyzing all individuals in the same group models allows for measurement of by-subject random effects of training and retention that can be directly related to one another through correlations[21]. Beyond allowing examination of the relationship between the effects of blocking on training and retention, extracting by-subject random effects allows investigation of the potentially interesting individual differences that may arise due to varied deficits. The same method was also used to look at the relationship between effects of blocking on training and generalization and the individual differences in the effects on generalization.

The analyses reported throughout this chapter were conducted using multilevel mixed models with random effects in R version 3.2.4 with the lme4 package (version 1.1-12). In all models, logistic regression was used to analyze letter accuracy data (percent of letters in a word produced correctly weighted by length of the word in letters). Although sessions were videotaped, response time was not analyzed in this study.

Training

[20] Note that individual models for each participant were also created, although they are not reported here. Although few results were significant at the individual level, the directions of effects were generally consistent with those found in the group model. Looking at individual differences from the group analyses revealed differences in the pattern of results for participant KSR2 as compared to the other participants, which is discussed at a later point in the main text. This different pattern was also observed in the individual participant analyses.

[21] It would be less appropriate to correlate fixed effects calculated in separate models of each individual because these measures may overestimate individual differences. Taking into account any group effects that may exist for this population by entering all data into the same model reduces this potential confound.

The first analyses concerned data collected during training. Does the trajectory of re-learning differ for words trained in semantic and segmental blocks as opposed to unrelated contexts? According to the first e-ILM predictions, semantic blocking should lead to interference relative to training in unrelated contexts, as should segmental blocking. That is, participants should show less improvement in accuracy during training for items trained in semantic vs. unrelated blocks and for items trained in segmental vs. unrelated blocks.

Model structure. Separate models were constructed to directly compare each blocking context to the unrelated context since the predictions of interest concern training in semantic blocks vs. unrelated contexts and segmental blocks vs. unrelated contexts. Both training accuracy models included block type (semantic, segmental, or unrelated context), training attempt within session (1-4), training session (1-4), two- and three-way interactions between those, and the control variables of word frequency, word length, days since last training session, and number of training trials since the target was last trained as fixed effects. A full random structure was implemented in each model, with random intercepts for subjects and items and a full random slope structure matching the fixed effect structure, excluding random slopes over items for block type, frequency, and length since each item had only one value for these variables.

Continuous variables (training attempt within session, training session, word frequency, word length, days since last training session, and number of training trials since the target was last trained) were centered and scaled. Block type was contrast coded in the separate semantic and segmental models. Data from the segmental context was excluded from the semantic model; block type was coded as semantic=1, unrelated=-

1. Likewise, data from the semantic context was excluded from the segmental model; block type was coded as segmental=1, unrelated=-1.

Explanation of included variables. As in the analyses of training for neurotypical participants presented in the previous chapter, both training attempt within session and session were included as variables as opposed to training attempt number across sessions. Including training attempt within session focused the analysis on effects during each training session, reducing the confounding effects of retention across sessions. A positive main effect of training attempt within session would show that participants improved as they practiced naming the same item repeatedly in a session. A positive main effect of session would show that participants improved as they completed additional training. The interaction between training attempt within session and session investigates whether the improvement within a session changed over the course of training sessions. For instance, if participants know most of the words by the beginning of the final session, they may show a smaller effect of training attempt within session, which would result in a negative two-way interaction.

The primary question of interest in this analysis was whether training in a blocked context increases difficulty during training. The most direct way to investigate this question is to look at the two-way interaction between block type and training attempt within session. If both types of blocking increase difficulty as predicted by the extension of the incremental learning model and as occurred in the previous experiment with neurotypical individuals, a negative interaction is expected in both the semantic and the segmental models whereby blocked items improve less within session than unrelated items.

The three-way interaction of block type, training attempt within session, and session is used to examine whether this effect changes over time as participants continue training.

Frequency and length of the words being trained were included to control for psycholinguistic factors that are known to affect the performance of these individuals. Days since the last training session and number of training trials since the target was last trained were included to evaluate effects of the spacing (in time and amount of intervening information) of training.

Results of training analysis. Figure 30 shows the participant performance during the training task. Tables F1 and F2 (in Appendix F) report the results of the analyses of this task.

Figure 30. Results of the training task over sessions from the dysgraphia treatment study.

The top left panel shows accuracy across the four training attempts within each session, collapsed across all four training session and all seven participants. Different colored lines represent the different training contexts whereby items in training blocks were segmentally related, semantically related, or unrelated.

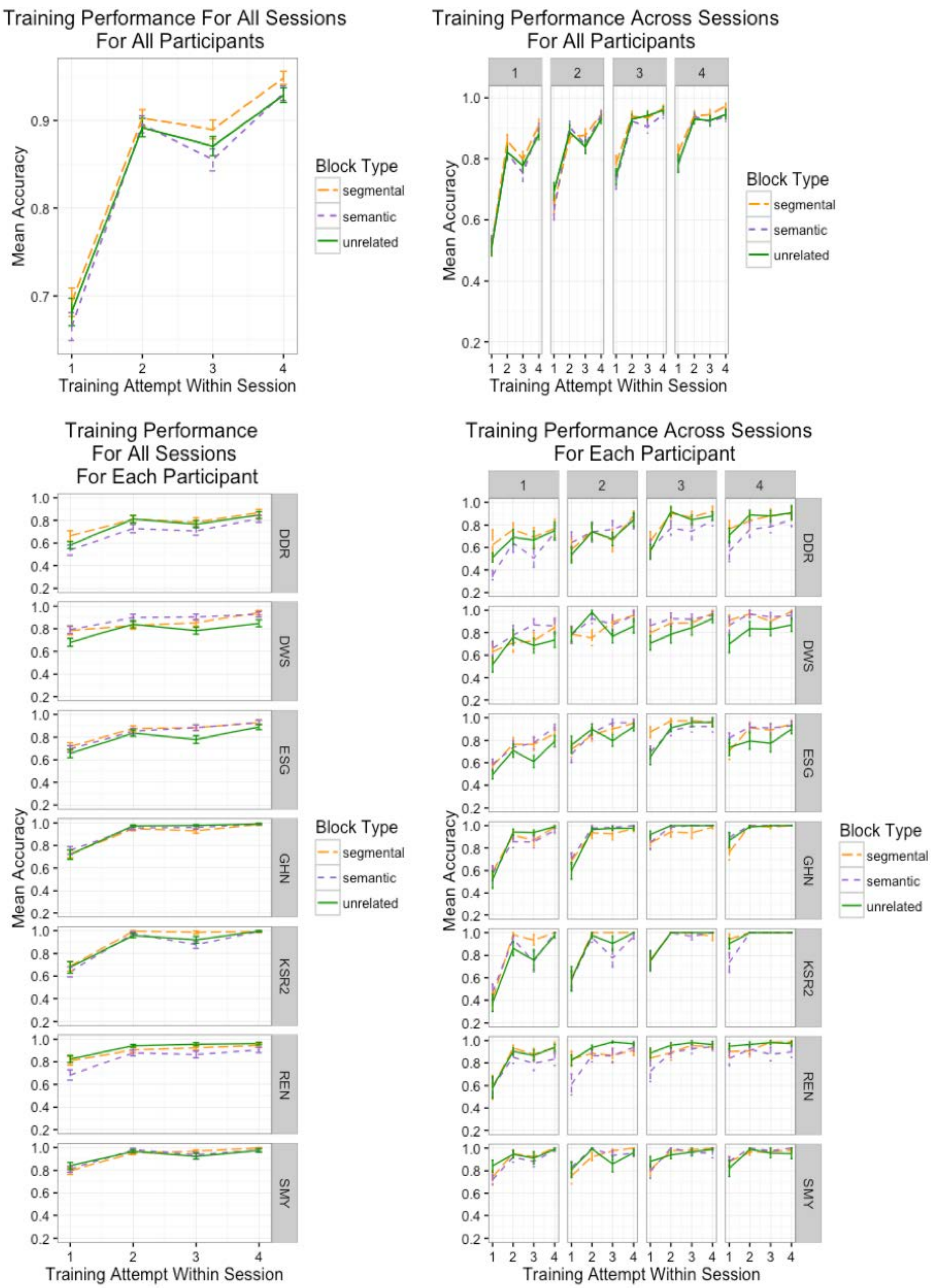
The top right panel shows accuracy across the four training attempts within each session for each of the four training sessions for the three training contexts, collapsed across all seven participants.

The bottom left panel shows accuracy across the four training attempts within each session for the three training contexts, collapsed across all four training session, for each of the seven individual participants.

The bottom right panel shows accuracy across the four training attempts within each session for each of the four training sessions for the three training contexts, for each of the seven individual participants.

The dependent variable in all figures is mean accuracy, measured as proportion of letters spelled correctly in each word.

Error bars represent one standard error of the mean, corrected for repeated measures.



Training Performance For All Sessions For All Participants
Training Performance Across Sessions For All Participants
Training Performance For All Sessions For Each Participant
Training Performance Across Sessions For Each Participant
Mean Accuracy
Training Attempt Within Session
Block Type
segmental
semantic
unrelated
DDR
DWS
ESG
GHN
KSR2
REN
SMY

The critical effect for evaluating the prediction that blocking leads to interference during training relative to training in unrelated contexts is the interaction of block type and training attempt within session. Although the predictions of the incremental learning model and the results of the previous experiment suggested that semantic blocking should increase training difficulty, no significant interaction was found in the semantic model (z=0.05, p=.963). However, it could be the case that semantic blocking increased difficulty for some of the individuals who participated, but that different deficits masked the effect. This possibility will be explored in later analyses looking at individual differences attempting to characterize the relationship between training and retention.

The e-ILM's predictions and the results of the previous experiment also indicated that segmental blocking should increase training difficulty. The segmental model, like the semantic model, did not provide support for this. In direct opposition to the prediction and past results, there was instead a significant positive interaction of block type and training attempt within session in the segmental model (z=2.69, p=.007), indicating a greater increase in accuracy over training attempts for items trained in the segmental as opposed to unrelated context across all seven participants. Again, it is possible that some individuals show the predicted effect, and this possibility will be investigated in later analysis.

Turning to the other effects of block type, the semantic model showed no main effect of block type, indicating that items trained in the semantic context were not produced more or less accurately than items trained in the unrelated context. However, there was a significant main effect in the segmental model: items trained in the segmental context were produced more accurately than items trained in the unrelated context

(z=3.54, p<.001). Although this effect does not directly show that there is greater improvement as a result of additional practice in segmental blocks as the interaction of block type and training attempt within session discussed above would, it does provide additional indirect support for the finding that the segmental context reduces training difficulty since these items are produced more accurately than those trained in unrelated contexts throughout the training task. There were no significant two-way interactions between block type and session, nor were there significant three-way interactions between block type, training attempt within session, and session, indicating that the main effects of block type and the effects of block type on improvement within session do not change as more training sessions are completed.

Other significant effects were revealed in the analyses of training data. There was clear evidence that participants learned over the course of the experiment, regardless of block type. There were consistent positive main effects of session, indicating that accuracy increased as additional training sessions were completed (z=7.32, p<.001 for the semantic model; z=7.40, p<.001 for the segmental model). Additionally, there were consistent positive main effects of training attempt within session, indicating that accuracy increased as the same word was practiced multiple times within the same training session (z=5.86, p<.001 for the semantic model; z=6.36, p<.001 for the segmental model). There was a positive interaction between these two effects in the semantic model (z=2.13, p=.033), indicating that the increase in accuracy within a session was larger for later sessions, consistent with the idea that participants were showing greater improvement within session as training proceeded, although this interaction was not significant in the segmental model.

The psycholinguistic variables of frequency and length were included to control for the length and frequency effects observed to affect accuracy for a number of the participants during screening. In the analyses, a significant effect of frequency indicating greater accuracy with higher frequency words was found in the segmental model (z=2.02, p=.043), although not in the semantic model. Significant effects of length indicating greater accuracy with shorter words were found in the semantic model (z=-3.93, p<.001), although not in the segmental model.

The variables of days since the last training session and trials since last trained were included to investigate the effects of spacing on training. There was not a significant main effect of days in either the semantic or the segmental model. There was a significant main effect of trials since last trained in both models (z=-2.11, p=.035 for the semantic model; z=-2.59, p=.010 for the segmental model), indicating that participants were more accurate when fewer trials intervened between attempts to name the same item. One contributor to this effect is likely that items were immediately repeated such that the second and fourth attempts directly followed the first and third attempts and probably benefitted from repetition priming.

Discussion of training analysis. Overall, the results of the training analysis from this study of individuals with dysgraphia relearning spellings were not consistent with the hypotheses of the e-ILM or the results of the previous study of neurotypical adults learning new words which predicted interference during training of items in both semantic and segmental blocks as opposed to unrelated contexts. While participants did improve both within and across sessions in all training contexts, there was a larger improvement within sessions for items trained in the segmental vs. unrelated context.

This facilitation effect directly contrasts with the expected interference effect. Semantic blocking did not lead to observable effects. As a whole, this group shows a different impact of training in blocks than predicted or previously demonstrated in neurotypical individuals. Because the other e-ILM predictions assume that blocking creates interference during training, this result indicates that the other prediction may not hold for this population either.

Assessment: Learning and Retention

The next goal of the study was to investigate the effects of blocking on retention, and to begin to evaluate whether either type of blocking was a desirable difficulty for this group of individuals with dysgraphia, testing the third and fourth e-ILM predictions. These are the predictions that connect the effects of blocking on training, derived from applying the incremental learning model of production of known words to learning of new words, to effects on long-term learning, suggested by other literatures. How do effects of blocking on training relate to learning outcomes? Did the participants remember the words that were trained in blocked contexts better or worse than those trained in unrelated contexts?

The e-ILM predicts a long-term learning advantage for items trained in semantic blocks, but a disadvantage for items trained in segmental blocks. However, contrary to the e-ILM predictions that both types of blocking should increase training difficulty, the training analysis revealed no effect of semantic blocking and a positive effect of segmental blocking. Since these direct predictions of the model were not upheld, the further predictions about blocking's effects on retention may not be upheld either. Alternatively, consider the prediction that semantic blocking is thought to be a desirable difficulty: to the extent that it increases difficulty in training, it increases long-term learning. If this prediction is accurate, there might not be an overall effect in the retention data since there was no overall effect in the training data. There could, however, be a relationship between individual effect sizes, which will be investigated later.

The e-ILM predicted that segmental blocking would reduce distinctiveness and thus be harmful for both training and retention. It had the opposite effect on training; it may also have an opposite effect that leads to a retention advantage over training in unrelated contexts. This would go along with the prediction that, in contrast to semantic blocking, segmental blocking is not a desirable difficulty. If this prediction is incorrect and segmental blocking is a desirable difficulty for this group of individuals with dysgraphia, an inverse relationship between training and retention might be observed here, with the advantage observed for segmental blocking in training reversing to become a disadvantage in retention.

The analyses of performance on assessments were designed to address not only the question of how blocking affects retention but also to address the efficacy of the training protocol. Are participants more accurate in producing words after training as compared to before? Is there evidence of learning relative to pre-training performance both immediately after training and at later follow-up time points? Does performance decline between the time-points immediately and later after training?

Model structure. A set of multilevel mixed models with random effects was constructed to analyze letter accuracy data collected during the assessment components of the treatment study. As in the previous analyses of training data, separate models compared items training in semantic blocks vs. unrelated contexts and segmental blocks vs. unrelated contexts. All models included block type (semantic, segmental, or unrelated context), time (pre-training, immediately post-training, or at follow-up), the two-way interaction between them, and the control variables of word frequency and word length as fixed effects. Days since training was not included as this was collinear with the categorically defined times (e.g., the time point immediately after training was always fewer days after training than follow-up). A full random structure was implemented in each model, with random intercepts for subjects and items and a full random slope structure matching the fixed effect structure, excluding random slopes over items for block type, frequency, and length since each item has only one value for these variables.

The continuous variables of frequency and length were centered and scaled. In order to evaluate all contrasts of the categorical variables of block type and time, a total of six models were constructed. In terms of block type, the semantic and unrelated contexts were compared to each other in models that excluded data from the segmental context; block type was coded as semantic=1, unrelated=-1. Correspondingly, the segmental and unrelated contexts were compared to each other in models that excluded data from the semantic context; block type was coded as segmental=1, unrelated=-1. In terms of time, two models included all three time points. These time points were Helmert coded. First, performance before and after training were compared to each other with pre-training=-1, immediately post-training=0.5, and follow-up as 0.5. Next, the time points after training were compared to one another, with immediately post-training as -1 and follow-up as 1. Two separate post-hoc models directly compared performance before training to performance directly after training that excluded data from follow-up;

time was coded as pre-training=-1, immediately post-training=1. Finally, an additional two post-hoc models directly compared performance before training to performance at follow-up that excluded data collected immediately after training; time was coded as pre-training=-1, follow-up=1.

Note that time was defined relative to training of each particular set of items (Table 11). Pre-training included all assessments of an item before it was trained, not just the initial administration before set 1 was trained. This inclusion of more data increased power and reduced potential confounds of effects of training other words on items that had not yet been trained. The immediately post-training time point included only the assessment that directly followed training of the set of which the item was a part. Follow-up included all assessments that occurred after the immediately post-training time point, not just those collected at the final assessment 6-8 weeks after the rest of treatment.[22] For example, for the post-set 2 assessment, the time point for items trained in set 1 was follow-up since they had previously been trained and assessed immediately after training; the time point for items trained in set 2 was immediately post-training since they had just been trained; and the time point for items trained in set 3 was pre-training since they had not yet been trained.

[22] The pattern of results was the same regardless of whether follow-up included the data collected later during training after the immediately post-training time point or was limited to the assessment 6-8 weeks later. Including this data increased the power of the experiment and allows more comprehensive tracking of retention of items trained in earlier sets. It was reasonable since the range of days since training overlapped. The later administrations during training occurred 22-119 days after training of an item; the follow-up administration occurred 56-189 days after training of an item.

Table 11. Definition of time points used in analyses. The left column lists the assessments that were conducted over the course of the study in chronological order. The right three columns separate the three sets of items trained in the experiment. These columns list the defined time point for data from each assessment, which differ based on when each set was trained relative to each assessment.

	Training Set		
Assessment	Set 1	Set 2	Set 3
Pre-treatment assessment	pre-training	pre-training	pre-training
Post-set 1 assessment	immediately post-training	pre-training	pre-training
Post-set 2 assessment	follow-up	immediately post-training	pre-training
Post-set 3 assessment	follow-up	follow-up	immediately post-training
Follow-up assessment	follow-up	follow-up	follow-up

Explanation of included variables. In this analysis, the effect of time was used to assess learning. Although there was evidence in the previous analysis that participants were more accurate in producing the spellings as more training sessions were completed, additional evidence of learning as a result of training can be drawn from this analysis and long-term retention can be investigated. In this analysis, comparing pre-training performance to performance after training (immediately post-training, follow-up, or the two combined) indicates whether there is learning at that time point as a result of the training, which would be shown by a positive main effect. Comparing performance immediately post-training and at follow-up evaluates whether this information is retained. A null effect would be consistent with retention of training gains, while a negative effect would indicate loss of information over time without practice.

The effect of block type evaluates whether words trained in different contexts are produced more or less accurately. A main effect in the models including pre-training data would be consistent with differences that arose before training.

The critical effect of interest in this analysis is the interaction between block type and time. This indicates whether training in a particular context compared to other contexts leads to different changes in performance over time. Of particular interest is the interaction in the models comparing performance immediately post-training and at follow-up. This interaction indicates whether training leads to better or worse retention. It most directly allows for evaluation of the e-ILM predictions. A positive interaction would indicate that a training in a particular context leads to increased retention, while a negative interaction would indicate that training in that context leads to reduced retention.

Frequency and length of the words being trained were also included to control for psycholinguistic factors that are known to affect the performance of these individuals.

Results of assessment analysis. Figure 31 shows participant performance during the assessment task. Tables F3-F8 (in Appendix F) report the results of the analyses of this task.

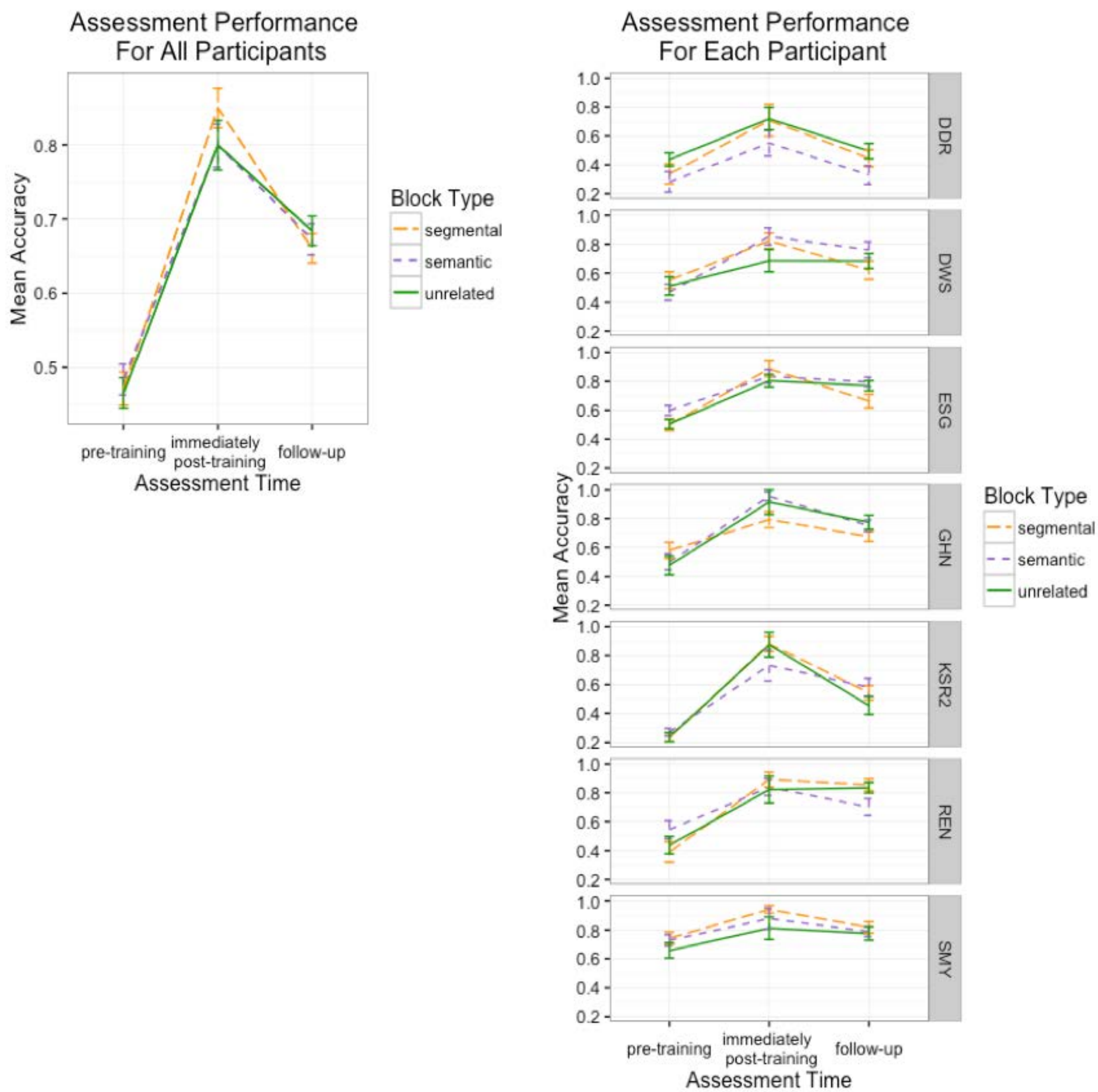


Figure 31. Results of the assessment task in the dysgraphia treatment study.

The left panel shows mean accuracy, measured as proportion of letters spelled correctly in each word, for all seven participants as they completed the assessment task of naming all 36 trained items before words were trained, immediately after they were trained, and at follow-up. Different colored lines represent the different training contexts whereby items in training blocks were segmentally related, semantically related, or unrelated. Error bars represent one standard error of the mean, corrected for repeated measures. The right panel shows mean accuracy on the assessment task for each of the seven individual participants.

The critical effect of interest in this analysis was the interaction of block type and time, which allows evaluation of whether the type of training affects learning and retention. Is the change over time different depending on the context of training? Is there more or less gain from pre-training to immediately post-training and/or follow-up or loss of information from immediately post-training to follow-up depending on whether words are trained in semantically related, segmentally related, or unrelated contexts? Results showed that there were no significant interactions between block type and time when block type contrasted the semantic and the unrelated contexts for the group as a whole (z=-0.14, p=.887 for before vs. after training; z=-0.60, p=.548 for immediately post-training vs. follow-up; z=0.43, p=.665 for pre-training vs. immediately post-training; z=-0.79, p=.430 for pre-training vs. follow-up). Just as in the training data analyses, some subset of the participants may experience the expected pattern of results whereby training in semantic blocks increases retention, but this is masked when all participants' data are combined. Further investigation of this possibility will be presented in a later section.

On the other hand, there were significant interactions between block type and time when block type contrasted the segmental and unrelated contexts. Participants achieved greater gains in accuracy between pre-training and immediately post-training for the items trained in segmental as opposed to unrelated blocks (z=2.22, p=.026). This is consistent with the finding in the previous analyses that segmental blocking reduced training difficulty. In contrast, there was greater loss in accuracy between immediately post-training and follow-up for the items from the segmental as opposed to unrelated blocks (z=-2.15, p=.032). Comparing all assessments before and after training or comparing pre-training to follow-up collapses across these opposing effects: there are not significant interactions of segmental vs. unrelated block type with time in those models (z=1.40, p=.160 for before vs. after training; z=-0.15, p=.882 for pre-training vs. follow-up). (There was a trend for greater improvement of items trained in segmental vs. unrelated contexts when comparing assessments before and after training.) The negative effect of training in segmental vs. unrelated blocks on retention after the completion of training is consistent with the prediction of the extended incremental learning account. However, these findings of opposite directions of effects of segmental blocking on learning as a result of training and on retention are not consistent with the extended incremental learning's prediction that segmental blocking is not a desirable difficulty. Rather, they align with the general idea of desirable difficulties: manipulations that increase training difficulty increase retention and vice versa.

Turning to other effects of the training context, there were no significant main effects of training in semantic blocks vs. unrelated contexts. However, there were significant effects in the segmental models. Participants were more accurate in writing words that were trained in the segmental as opposed to unrelated context in the models comparing all three time points (z=2.04, p=.041) and comparing pre-training and immediately post-training (z=2.16, p=.031), although this did not reach significance in the model comparing pre-training and follow-up (z=0.74, p=.458). This overall advantage for words that were trained in segmental vs. unrelated contexts may have arisen even before training commenced, as shown by the main effects of block type that were observed even when pre-training data was included. This is possible since items were selected to fall within an accuracy range at baseline but groups were not matched for accuracy and since performance could change between baseline testing and the pre-treatment assessment.

Other effects observed in the analysis of the assessment data concern the effectiveness of the treatment administered in this study. Participants did learn the trained items. Positive main effects of time demonstrate that they were more accurate in producing items after training than before training regardless of training context. This was true when pre-training was compared to both post-training time points combined (z=9.18, p<.001 for the semantic model; z=6.65, p<.001 for the segmental model), when pre-training was compared only to immediately post training (z=7.38, p<.001 for the semantic model; z=6.93, p<.001 for the segmental model), and when pre-training was compared only to follow-up (z=7.30, p<.001 for the semantic model; z=4.63, p<.001 for the segmental model). While participants did maintain learning to the extent that follow-up performance was better than pre-training performance, there was significant loss of information between the assessments administered immediately after training and at follow-up (z=-4.20, p<.001 for the semantic model; z=-4.15, p<.001 for the segmental model). The treatment protocol was successful on the whole, even though there was some forgetting when practice ended.

Psycholinguistic variables did not have significant effects on performance. The effect of length was consistently negative (i.e., worse performance for longer words), although this did not attain significance in any of the models. The effect of frequency was consistently positive (i.e., better performance for more frequent words) in all models except the semantic model comparing performance before training to performance at follow-up, although it was not significant in any of the models either. These failures to find effects may be due to restricted range: the sets were matched on length and frequency when the stimuli were selected.

Discussion of assessment analysis. The results of the assessment analyses were not consistent with the e-ILM predictions or the results of the previous study of neurotypical adults with regard to training in semantic vs. unrelated blocks. Here, there was a null effect in contrast to the expected advantage for the spellings of items trained in semantic blocks as opposed to unrelated contexts. The failure to find significant interactions for the semantic vs. unrelated context with time is consistent with the null finding in the training analysis. In analyses at the group level, semantic blocking does not appear to be a desirable difficulty. Semantic blocking neither reduced nor increased training difficulty, and it neither reduced nor increased retention across all participants.

Segmental blocking, however, did have a significant negative effect on retention of spellings. This is consistent with the prediction of the extended incremental learning account, although not consistent with the results of the previous study with neurotypical learners in that no significant effect was found in the previous study. In contrast to the prediction of the e-ILM, however, segmental blocking fits the profile of a desirable difficulty with a direct relationship between training difficulty and retention. Note that the direction of the effects for the group of individuals in this study was such that segmental blocking reduced training difficulty and long-term retention as compared to training in an unrelated context. The finding from the present study is in line with the results of Storkel, Bontempo, and Pak (2014), who found that form similarity between words being learned increased immediate performance but reduced recall at a later time.

The effects discussed here are at the group level, collapsing across all seven participants. The next analysis looks more directly at the relationship between training difficulty and retention to see if these effects are related at the individual level, providing further evaluation of whether either type of blocking is a desirable difficulty.

Individual Differences: Correlations between Training and Retention

The next goal of the analyses was to more directly evaluate the fourth e-ILM predictions regarding whether either semantic or segmental blocking during training was a desirable difficulty. The model predicts that semantic blocking should be a desirable difficulty that increases difficulty during training and increases retention, but that segmental blocking should not be a desirable difficulty. The results of the previous study of neurotypical individuals learning new words were generally consistent with a broad construal of these predictions in that there was some evidence that semantic blocking served as a desirable difficulty and no clear evidence that segmental blocking did so. (Note, however, that segmental blocking did not have the predicted opposite relationship whereby increased difficulty during training led to reduced retention either.) In order to assess whether these predictions about desirable difficulty are supported by the data of the present study regarding individuals with dysgraphia relearning spellings, two types of information can be considered. These analyses are parallel to those used in the previous study. First, the pattern of results across subjects can be evaluated. This was discussed to some extent in the previous section of this chapter. Second, correlations between the by-subject random effects of blocking on training and retention can be calculated to more directly look at the relationship between training and retention at the individual level. Effects that are masked by grouping all participants together may be uncovered by looking at the pattern of individual differences. Looking at individual effects also provides an opportunity to speculate about possible relationships between deficits and the effects of blocking.

Pattern of results across participants. The previous analyses of the data collected during the training and assessment phases of the dysgraphia treatment study showed that there were no significant effects of training in semantic blocks as opposed to unrelated blocks for the group as a whole on either training or retention. Therefore, there was no clear relationship between semantic blocking effects on training and retention, and semantic blocking did not fit the profile of a desirable difficulty. This does not support the prediction of the e-ILM, and it differs from the results of the previous study.

The situation for segmental blocking is different than that of semantic blocking. Training in segmental blocks relative to training in unrelated blocks reduced training difficulty and led to reduced retention of the spellings of the items from the time point immediately after training to follow-up for the group as a whole. This fits the profile of a desirable difficulty in that training difficulty and retention are related. However, this manipulation reduced difficulty and retention instead of increasing them: for this group of participants, segmental blocking neither created difficulty nor had desirable long-term effects. This result contradicts the predictions of the e-ILM and the results of the previous study. However, it fits with the general theory of desirable difficulties: manipulations that increase or reduce training difficulty correspondingly increase or reduce retention.

Examining individual differences. The individuals with dysgraphia who participated in the study were not a homogeneous group: they exhibited a range of different cognitive deficits. Analyzing individual differences may reveal effects that do not appear in the analysis of the group as a whole.

The e-ILM predictions, along with the results of the previously described experiment with neurotypical participants learning new vocabulary, suggest that semantic blocking may be a desirable difficulty, increasing retention to the extent that training difficulty is increased. Although this effect was not found for the group as a whole, the model and past results lead to the prediction that individual effects of semantic blocking on training (as measured by the two-way interaction between semantic vs. unrelated block type and training attempt within session) should be negatively correlated with individual effects of semantic blocking on retention (as measured by the two-way interaction between semantic vs. unrelated block type and immediately post-training vs. follow-up assessment time).

In contrast, the e-ILM predicts that segmental blocking is not a desirable difficulty, a prediction that was broadly consistent with the findings of the previous study. If this is the case, there should not be a significant negative correlation between training and retention effects since there are not opposite effects on training and retention. There were, however, more specific predictions about the relationship between the effects of segmental blocking on training and retention. According to the e-ILM, application of the incremental learning model of production to the training of new words leads to the predictions that training in segmental blocks vs. unrelated contexts should increase training difficulty and reduce distinctiveness. According to the assumptions linking training and long-term learning, this reduced distinctiveness should lead to reduced retention for the items trained in segmental blocks vs. unrelated contexts. This means that segmental blocking should be the opposite of a desirable difficulty, with increased difficulty during training leading to reduced retention. That is, a positive correlation between training and retention effects would be expected since there the e-ILM predicts a negative effect on training and a negative effect on retention. On the other hand, the empirical results across participants already observed in this study suggests that segmental blocking may fit the pattern of a desirable difficulty with opposite effects on training and retention: training in segmental blocks as opposed to unrelated contexts reduced training difficulty and reduced retention at the group level. There may therefore be a negative correlation between training and retention effects, indicating an inverse relationship between training and retention at the individual level that is consistent with the idea that segmental blocking fits the profile of desirable difficulty.

Correlating blocking effects. In order to test these predictions, the random effects for each participant were extracted from the group models of training and retention. This analysis followed the suggestions of Mirman (2014) that were also used in the analysis of individual differences in the previous experiment regarding the effects of blocking on learning of new words by neurotypical participants. In the present analysis, the by-subject random effects of block type by training attempt within session were extracted from the training models and the by-subject random effects of block type by immediately after training vs. follow-up time were extracted from the assessment models. These by-subject random effects quantify each participant's deviation from the overall pattern predicted by the fixed effects. That is, they show much each individual's effect differs from the group effect. By calculating the correlations between these random effects, it is possible to examine the relationship between training and retention. For example, it is possible to see if those who experience a larger effect of one type of blocking on training also show a larger effect of that type of blocking on retention and if these effects are in the same direction. Pearson's correlations were calculated between random effects with the same block type contrast (e.g., semantic vs. unrelated training was correlated with semantic vs. unrelated retention).

Results of correlation analysis. Figure 32 shows the correlations between by-subject random effects of training and retention.

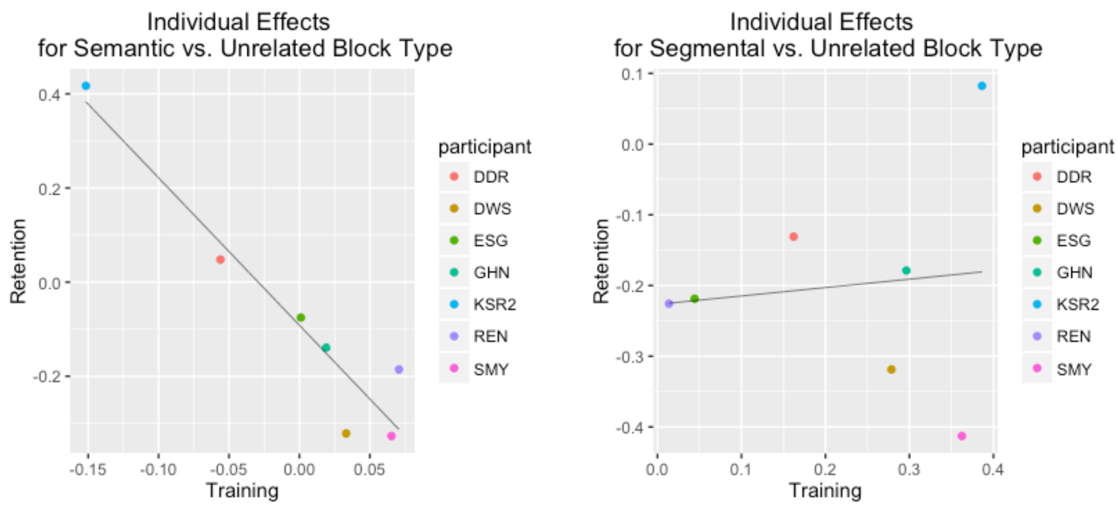


Figure 32. Relationship between the by-subject random effects of training and retention.

For both panels, by-subject random effects are calculated as the sum of the fixed effect for the group model and each individual's adjustment, which was extracted from the random effects. The x-axis for both panels is the training effect, measured by the interaction of block type and training attempt within session from the training data. The y-axis for both panels is the retention effect, measured by the interaction of block type with time point, which contrasted immediately post-training vs. follow-up. Colored points represent each participant's values. The mapping between participant identity and color is the same throughout the figure. The black lines are linear regression lines that depict the correlation trend.

Each panel depicts the individual effects of training and retention when a different block type contrast is used. In the left panel, block type contrasts semantic and unrelated contexts. In the right panel, block type contrasts segmental and unrelated contexts.

The critical effects for evaluating the predictions about semantic and segmental blocking as desirable difficulties were the correlations between individual effects of blocking on training and retention. Separate comparisons of semantic vs. unrelated block type and segmental vs. unrelated block type were made. When block type compared the semantic and unrelated contexts, there was a significant negative correlation between the effects of semantic blocking on training and retention(r=-.95, p=.001).[23] This negative correlation is consistent with the idea that semantic blocking is a desirable difficulty. Although there were not consistent effects across the group as a whole, examining the relationship between individual differences showed that the individuals who experienced the greatest difficulty as a result of training in semantic blocks as opposed to an unrelated context were the same individuals who demonstrated the largest benefit in retention after training and vice versa. This aligns with the predictions of the incremental learning model and the results of the previous experiment with neurotypical participants.

However, when block type compared the segmental and unrelated contexts, there was not a significant correlation (r=.11, p=.808). Although the overall pattern in the group analysis suggested that segmental blocking was a desirable difficulty, with a significant positive group effect for training and a significant negative group effect for retention, the individual participants did not demonstrate a clear relationship. The ones who experienced the greatest training difficulty were not the same ones who demonstrated the greatest retention benefit. This null effect weakens the claim that

[23] As in the previous study with neurotypical participants, model-based Monte Carlo simulations were constructed to compare this correlation to chance. Models were fit to simulated data 100 times, and the random effects were extracted and used to calculate correlations in the same way as was done for the real data. Results showed that the real correlation between training and retention effects on accuracy for the semantic vs. unrelated training contexts was more negative than all correlations from the simulations, suggesting that the real correlation was significant at the $p<.01$ level.

increases in training difficulty due to training in segmental blocks is directly related to increased retention for those items, even though the group effect is consistent with the claim that training in segmental blocks reduces training difficulty and retention.

Calculating individual effects. Looking at the patterns of the individual effects, beyond assessing their statistical significance, can be informative as well. Examining which participants have positive and negative effects of training and retention for the various block types could potentially provide information about how different deficits lead to different block type effects. Individual effects were calculated by summing the fixed effect from the group model with the individual adjustment extracted from the by-subject random effects (following similar methods to Bonin, Méot, Lagarrigue, & Roux, 2015). The individual adjustment alone is informative with regards to whether the individual's effect is more positive or more negative than the group effect. Adding the constant fixed effect provides information about the direction of the effect for the individual. For example, if the fixed effect is large and positive, the largest negative individual adjustment may still represent a positive effect at the individual level.

Results of individual effect analysis. Individual effects are shown in Table 12. These data match what is presented in the graph of Figure 32.

Table 12. By-subject random effects, calculated as the sum of the fixed effect for the group and each individual's adjustment.

	training effects: block type * trial within session		retention effects: block type * immediately post-training vs. follow-up time	
participant	semantic vs. unrelated	segmental vs. unrelated	semantic vs. unrelated	segmental vs. unrelated
DDR	-0.056	0.162	0.048	-0.131
DWS	0.033	0.279	-0.322	-0.319
ESG	0.001	0.044	-0.075	-0.219
GHN	0.019	0.296	-0.139	-0.179
KSR2	-0.152	0.386	0.417	0.082
REN	0.071	0.014	-0.186	-0.226
SMY	0.065	0.362	-0.327	-0.413

Individual effects of semantic blocking. Looking at the effects of semantic vs. unrelated block type on training and retention, five of the seven participants had positive effects during training. This indicates that they improved more quickly as they practiced the same item within a session when that item belonged to a semantic block as opposed to an unrelated block, consistent with training in semantic blocks reducing training difficulty. Participants DDR and KSR2 showed negative effects, indicating that for these individuals, training in semantic blocks increased training difficulty. The single largest effect was the negative effect experienced by KSR2 (-0.152), which had more than twice the magnitude of the largest positive effect experienced by REN (+0.071). DDR's negative effect (-0.056) was also relatively large in magnitude, with an absolute value larger than three of the five positive effects. Moving to the retention effects, the same five individuals who experienced positive effects in training showed negative effects in retention. Likewise, DDR and KSR2 showed positive effects in retention in contrast to their negative effects in training. Again, KSR2's effect had the largest magnitude. As expected from the correlation analysis, increased difficulty in training leads to superior retention and vice versa.

Connecting individual effects of semantic blocking to deficits. Given this data, it is possible to speculate about the way different deficits impact the effect of semantic blocking. While it is difficult to say anything definitive about DDR's deficits since his responses on screening assessments were so limited, KSR2's deficit does show several potentially interesting contrasts with the deficits of other participants. During screening, KSR2 did not show effects of frequency, unlike all other participants except DDR. Frequency effects are interpreted as indicating deficits affecting the orthographic long-term memory. All other participants except DDR and SMY, for whom there may not have been sufficient power to detect an effect, also showed effects of length that indicated deficits affecting orthographic working memory. KSR2 had trouble accessing the correct long-term orthographic memory representation from semantic information as evidenced by his many semantic errors but lack of length and frequency effects, indicating a deficit affecting lexical selection. The other participants (with the possible exception of DDR whose responses were too limited to accurately characterize) had deficits that made selection and maintenance of graphemes in working memory difficult, indicating deficits affecting segmental encoding. The individual effects reported here speculatively suggest that there might be a dissociation in the effects of semantic vs. unrelated blocking that depends on deficit. For KSR2 (and potentially DDR), who had the most severe deficits in lexical selection relative to segmental encoding, semantic blocking increased training difficulty and improved retention relative to training in unrelated contexts. For the other individuals, semantic blocking reduced training difficulty and reduced retention relative to training in unrelated contexts. Note, however, that it is not possible to definitively attribute the differences for KSR2 to his spelling deficit: he may be different from the other participants in additional important ways. Further study with other individuals with deficits similar to KSR2 is needed to confirm the speculation about the cause of different effects for this participant as opposed to the majority of the group.

Semantic blocking fit the pattern of a desirable difficulty across all participants, but it was not in fact desirable in terms of long-term outcome for all. Semantic blocking as opposed to training in unrelated contexts was beneficial in the long run in terms of retention for those with the most severe lexical selection vs. segmental encoding deficits, while it was detrimental for those with clearer segmental encoding deficits. Future studies with a larger number of individuals with more clearly contrasting deficits are needed to verify these speculations about the relationship between deficit type and the effects of semantic blocking.

Individual effects of segmental blocking. Looking at the effects of segmental vs. unrelated blocking, all seven participants showed positive effects during training, indicating that segmental blocking reduced training difficulty relative to training in unrelated contexts. This was reflected in the significant effect of segmental blocking found in the group analysis. Six of the seven participants had negative effects on retention of training in segmental blocks relative to unrelated contexts, indicating that training in segmental blocks was detrimental to long-term retention. Again, this was reflected in the significant effect found in the group analysis. Only one participant, KSR2, had a positive effect of segmental blocking relative to unrelated blocking in retention, and the magnitude of this effect was the smallest of all seven participants. Given that only one individual demonstrated this reversed effect, it is difficult to say whether this difference should be attributed to his spelling deficit or to other individual variation, although do note that this individual was the one with the most severe lexical selection relative to segmental encoding deficit as discussed above. Segmental blocking seems to have more uniform effects than semantic blocking, at least in this particular set of participants. Only one individual showed a long-term benefit from segmental blocking, though all seven showed short-term benefits during training. Further studies are needed to determine if these results extend to other individuals and to see if there are different effects depending on deficit type.

Discussion of individual differences analysis. The analysis of individual differences thus showed that semantic blocking is a desirable difficulty, with a significant negative correlation between the effects of training and retention. However, it is possible to speculate that whether participants experience the beneficial side of this correlation (i.e., increased training difficulty and retention relative to training in unrelated contexts) as opposed to the detrimental flip side (i.e., decreased training difficulty and retention relative to training in unrelated contexts) may depend on their deficit profiles. In this small sample of participants with complex deficits, it was those who seemed to have the most severe lexical selection vs. segmental encoding deficits that benefitted most in terms of retention, while those with segmental encoding deficits benefitted less. Segmental blocking also seems to inversely affect training and retention, although this was more consistently true across participants and there was not a significant correlation between the two. Only one participant, one of those who also benefitted from semantic blocking, showed positive retention effects. More investigation is needed to follow up on these speculations about the relationship between deficit type and effects of blocking.

These results show that, at a practical level, just because an effect can be characterized as a desirable difficulty with opposite effects on training and retention, does not mean that it should be uniformly applied in treatment across all individuals. Effects that are beneficial in neurotypical populations may not be beneficial when applied to individuals with cognitive deficits. For most of the individual participants in this study, both semantic and segmental blocking led to positive effects on training but negative effects on retention. That is, both types of blocking fit the pattern of a desirable difficulty, but they neither induced difficulty nor had desirable effects on retention for the majority of participants. The profiles of the individuals undergoing treatment need to be considered.

Generalization

The present study also provided the opportunity to investigate the clinically relevant question of whether there was generalization to untrained items as a result of this treatment protocol. Although item-specific improvement (i.e., increased accuracy only for trained words) would show that there are benefits of the treatment protocol, demonstrating generalization to untrained words would be an even better outcome at a practical level. To evaluate this question, spelling performance before and after treatment was compared using a generalization list consisting of items from the same length and frequency range as treated items that had first been spelled during screening and baseline assessment. Details of the generalization analysis are presented in Appendix G. To summarize the results, analysis showed that there was not significant generalization to untrained items: untrained items from the generalization list were spelled no more accurately after treatment than before treatment.

Analyses also allowed for investigation of whether generalization differed depending on the context of training. That is, were there different patterns of generalization for untrained words that are semantically related to those trained in semantic blocks or segmentally related to those trained in segmental blocks as compared to those that are unrelated to those trained in blocks? Finding differences would provide indirect support for the e-ILM predictions regarding distinctiveness. Training in semantic blocks was predicted to enhance distinctiveness by weakening connections between shared features and lexical nodes while strengthening connections between distinctive features and lexical nodes. This could have negative consequences for semantically related words that are not trained: the connections between shared features and their lexical nodes were repeatedly weakened by training of the related words and so contribute less to activation when the untrained word becomes a target than they would have if related words had not been trained. It is therefore predicted to be more difficult to produce an untrained word that is semantically related to items trained in semantic blocks after those semantically related items are trained. Unrelated, untrained items should not be affected in the same way since there is not repeated weakening of any of the connections between features and lexical nodes. Therefore, reduced generalization to untrained words that are semantically related vs. unrelated to items trained in semantic blocks is expected.

On the other hand, the e-ILM predicts that segmental blocking reduces distinctiveness by strengthening connections between lexical nodes and shared segments while weakening connections between lexical nodes and distinctive segments. This should have positive consequences for segmentally related words that are not trained: the connections between shared segments and lexical nodes are repeatedly strengthened by training of the related words and so can contribute more when the untrained word becomes a target than they would have if related words had not been trained. It might therefore be less difficult to produce an untrained word that is segmentally related to items trained in segmental blocks after those segmentally related items are trained. However, there could also be negative consequences for segmentally related words that are not trained: the connections between distinctive segments and non-target lexical nodes are repeatedly weakened by training of the related words and may contribute less when the untrained word becomes a target than they would have if related words had not been trained. It might therefore be more difficult to produce an untrained word that is segmentally related to items trained in segmental blocks after those segmentally related items are trained. The balance of these two effects makes it more difficult to make predictions about generalization to untrained words that are segmentally related to items trained in segmental blocks as compared to unrelated items. If the strengthening of connections between shared segments and lexical items matters more, there could be increased generalization; if the weakening of connections between distinctive segments and lexical items matters more, there could be reduced generalization; or the opposing forces of strengthening and weakening could cancel one another out, leading to no significant effects on generalization. Therefore the examination of segmental blocking's effects on generalization is more exploratory. There is a clear prediction that semantic blocking will lead to reduced generalization, but there is not a corresponding clear prediction about how segmental blocking affects generalization.

Results showed that there were no differences in generalization at the group level for untrained items that were semantically related vs. unrelated to items trained in semantic blocks or for items that were segmentally related vs. unrelated to items trained in segmental blocks. (Details of the analysis are presented in Appendix G). The null effect for semantic blocking is not in line with the e-ILM, which predicted that there would be reduced generalization for untrained items that were semantically related vs. unrelated to items trained in semantic blocks. With regards to the more exploratory analysis of the effects of segmental blocking on generalization, the results are consistent with the idea that the reduced distinctiveness caused by training in segmentally related vs. unrelated blocks does not affect generalization, possibly because the weakening of connections between distinctive segments and lexical nodes that might otherwise lead to reduced generalization is canceled out by the strengthening of connections between shared segments and lexical nodes that might otherwise lead to increased generalization.

In interpreting these results, it is important to remember that the predictions of the e-ILM were only indirectly addressed here. There are many potential reasons the expected effect for semantic blocking was not observed, including low power to detect differences between generalization conditions since there were not overall generalization effects and potential masking of individual effects in the group analysis. However, it is also possible that null effects were seen here for both blocking contexts because training in blocks did not have the expected effects for the majority of participants. For most, training in blocked vs. unrelated contexts did not lead to interference but rather induced facilitation. This type of training thus may not have had the expected effects on distinctiveness and therefore would not affect generalization. The results are not conclusive. Future studies should look more directly at the effects of training in blocks on distinctiveness to better evaluate the predictions of the e-ILM in rehabilitation contexts.

General Discussion

In this study, seven individuals with dysgraphia participated in a treatment protocol in which they were trained to spell target words that were arranged into semantically related, segmentally related, and unrelated blocks. This study extended the work presented in the previous chapter in which neurotypical participants learned new words in semantically related, segmentally related, and unrelated blocks. The dysgraphia treatment study allowed investigation of whether the same effects were shown in this different population as they performed the somewhat different task of relearning previously known words, providing an additional opportunity to test the predictions of the extensions of the incremental learning model first presented in Chapter 3.

The present study directly addressed the first set of e-ILM predictions that result from directly applying the incremental learning model of production of known words to training of unknown words, testing whether semantic and segmental blocking induce interference during training relative to training in unrelated contexts. The second set of predictions, regarding changes to distinctiveness as a result of training in blocks, was not directly evaluated in this study, as shared and distinctive features were not probed. However, the predicted consequences of distinctiveness were indirectly addressed in testing the third and fourth predictions that connect the effects of training in blocks that are derived from direct application of the model to effects on long-term outcomes that are derived from the learning literature. The third set of predictions addresses effects on retention after training in blocks, relying on the assumptions that training in blocks changes distinctiveness and that these changes in distinctiveness affect retention. The fourth set of predictions addresses the relationship between training difficulty and long-term learning outcomes, which may differ depending on whether difficulty during training leads to enhanced or reduced distinctiveness. Below, I consider how the results found in this study relate to the e-ILM predictions.

Semantic Blocking

First, consider the predictions regarding the effects of training in semantically related blocks.

Semantic prediction 1: Training in semantic blocks leads to interference during acquisition relative to training in unrelated contexts. This prediction was investigated by comparing the improvement for items trained in semantic and unrelated blocks as participants practiced producing each item more times within a training session. According to the prediction, participants should show reduced improvement in accuracy during training for items being trained in semantic blocks vs. items being trained in unrelated contexts. Results of the study showed that items trained in semantic blocks did not exhibit increased or reduced improvement within sessions relative to items trained in unrelated contexts at the group level of analysis. That is, the prediction of interference during acquisition was not supported for the group as a whole. In looking at individual effects, only two of the seven participants had negative effects during training consistent with increased difficulty due to training in semantic vs. unrelated blocks. This finding provides further evidence against the prediction, which did not hold for this group of individuals with dysgraphia. The failure to find consistent interference in training also goes against the results of the previous study of neurotypical adults learning new words, where the predicted interference was observed.

Semantic prediction 3: Training in semantic blocks is beneficial for retention relative to training in unrelated contexts, assuming that increased distinctiveness leads to better long-term learning outcomes. This prediction was examined by comparing accuracy for the spellings of items trained in semantic blocks vs. unrelated contexts on recall assessments administered immediately post-training and at follow-up several weeks later. According to the prediction, participants should show better retention of items trained in semantic blocks than of items trained in unrelated contexts. However, the results did not support this prediction: although participants showed learning relative to the pretraining assessment, there were no significant differences in the retention of items trained in semantic blocks as opposed to unrelated contexts between the post-training and follow-up assessments for the group as a whole. In looking at individual differences, only two participants showed the expected positive effect indicating better retention of items trained in semantic vs. unrelated contexts. These were the same individuals who showed interference during training for the items trained in semantic vs. unrelated contexts. This finding provides further evidence against the prediction, which did not hold for this group of individuals with dysgraphia. The failure to find a consistent retention advantage for the spellings of items trained in semantic vs. unrelated context also goes against the results of the previous study of neurotypical adults learning new words, where there was some support for the predicted retention advantage for items trained in semantic blocks as opposed to unrelated contexts.

Semantic prediction 4: Semantic blocking is a desirable difficulty: increased training difficulty (interference) leads to increased long-term retention due to increased distinctiveness. This prediction was assessed in two ways. First, the pattern of results across participants was considered. According to the prediction, participants as a whole should show interference during training but an advantage in retention for items trained in semantic blocks as opposed to unrelated contexts. As described above, there were not consistent effects of training in semantic blocks vs. unrelated contexts for either training or retention. The pattern of results in the group analysis was not consistent with the prediction that semantic blocking is a desirable difficulty. However, the participants in this study were a heterogeneous group with complex deficits. It was possible that, although there was not a consistent trend for the group as a whole, there might be a systematic relationship between the effects such that the participants who experienced the most interference during training also experienced the biggest retention advantage for words trained in semantic blocks vs. unrelated contexts. Finding such a relationship between training and retention effects would support the prediction that semantic blocking is a desirable difficulty. This possibility was examined by calculating the correlations between individual differences in semantic blocking effects on training and retention. Results showed a negative correlation: the participants who demonstrated the most interference during training due to training in semantic blocks vs. unrelated contexts also showed the largest advantage in retention for those items. This finding is consistent with the prediction: semantic blocking fits the profile of a desirable difficulty, even though it did not lead to increased interference during training and long term retention advantages as compared to training in unrelated contexts for all participants. This matches the results of the previous study, in which there was some evidence supporting the idea that semantic blocking is desirable difficulty for neurotypical adults learning new words.

Discussion of semantic blocking. The results of the study did not support the direct prediction of the e-ILM: training in semantic blocks did not increase training difficulty for the group as a whole relative to training in unrelated contexts. Given that this effect was not observed, it is unlikely that training in semantic blocks increased distinctiveness as predicted. In evaluating the prediction that connected training effects to long-term learning, it is thus unsurprising that training in semantic blocks did not lead to better retention than training in unrelated contexts for the group as whole. The prediction regarding retention relies on the assumption that increased retention is a result of increased distinctiveness, and the null result regarding training difficulty suggests that this increase may not have occurred in this group.

Although the results for the group as a whole did not align with the model's predictions, analysis at the individual level did reveal a relationship between training and retention effects in line with that predicted by the learning literature. To the extent that training in semantic blocks vs. unrelated contexts led to increased interference during training, it also led to long-term retention advantages for those items. That is, semantic blocking is a desirable difficulty, in line with e-ILM prediction. While the main effects for the group did not follow the predicted pattern, individual differences show that the predictions did hold for at least some individuals and supported claims about underlying mechanisms. Those who experienced increased difficulty during training as a result of practice in semantic blocks vs. unrelated contexts also experienced increased retention for those words trained in semantic blocks, potentially as a result of increased distinctiveness, while those who experienced reduced difficulty during training also experienced reduced retention, potentially because distinctiveness was not increased.

While semantic blocking fits the e-ILM prediction regarding its status as a desirable difficulty, further research is required to investigate the characteristics of those who benefit from training in semantic blocks and those who do not. As described above, many of the individual participants experienced effects in the opposite direction than predicted by the e-ILM. For five of the seven participants, training in semantic blocks vs. unrelated contexts led to increased improvement during training and reduced long-term retention. The general pattern of a desirable difficulty is present, but for these individuals the manipulation neither increased difficulty nor had "desirable" effects on long-term learning. Why might this be? The answer is likely related to the deficits experienced by these individuals. The individuals in this group have complex deficits that make them less than ideal for addressing the theoretical problems of how deficits relate to the behavioral outcomes of this study. Future research with individuals whose deficits are better suited to investigating the theoretical questions is obviously necessary. However, it is still possible to speculate about how different deficits may lead to the different patterns of results seen here

Although deficits were mixed to some extent for all the participants, some general patterns can be noted. Two participants, DDR and KSR2, experienced interference as a result of training in semantic vs. unrelated blocks and then showed long-term advantages for those items. While it is difficult to characterize DDR's deficit, KSR2 appeared to have the most severe problem affecting lexical selection relative to segmental encoding.

He made many semantic errors that were consistent with difficulty activating the correct lexical node from its semantic features, although he did not exhibit length or frequency effects that would be consistent with difficulty selecting and maintaining the segments of words. The other five who showed reversed training and retention effects appeared to have deficits affecting segmental encoding and orthographic working memory, with frequency and length effects indicative of difficulty selecting and maintaining the correct segments to produce the words. (Some also had deficits that affected lexical selection as well.)

To speculate, there are several possible ways that the different deficits may have led to the different effects of semantic blocking, which are not mutually exclusive. One possibility is that those who benefitted were re-learning these items as if they were entirely new words. If the semantic and/or lexical representations themselves were too damaged to be salvageable, new representations would need to be created and integrated into the production system. If the weights between the representations were too weak, new connections might need to be formed and the connection strengths adjusted, or the adjustments to the weakened connections might look like forming of new connections. If either of these is the case, these participants could use the same learning system that neurotypical participants use when they encounter new words. In such a situation, the same effects seen in the previous study would be expected; this is what was observed for the individual with relatively more severe deficit in lexical selection as opposed to segmental encoding.

Other possibilities can be applied to speculatively explain the reversed effect for the participants with segmental encoding deficits. One possibility concerns residual activation. As has been discussed earlier in this dissertation, activating a word also leads to activation of its related neighbors. When a target is selected, this activation does not immediately dissipate, but rather some of this activation along with activation for the target persists to the next trial. For example, in producing the target *cat,* the lexical node of *dog* and its segments receive some activation that may continue onto the next trial. This may influence performance on future trials; residual activation is likely one of the reasons that semantic blocking effects emerge in normal production. When naming *dog* after *cat,* residual activation for *cat* which remains after its selection makes *cat* a stronger competitor. This increased activation for *cat* leads to larger changes to its weights when *dog* is produced since weakening is proportional to the amount of activation for non-targets. This may accelerate the interference effects seen in blocked cyclic naming tasks. Note, however, that there is also weak residual activation for the new target *dog* as a result of its activation as a competitor on the previous trial. This residual activation would be helpful in selecting *dog*. One problem that the individuals with segmental encoding deficits may face is reduced activation of lexical nodes and segments. For example, they may not be able to activate *dog* and the letters D-O-G strongly enough to select them. In this case, residual activation may provide a boost that allows for selection of the target. In a semantically related block, in which the different items being trained are related, residual activation from items that competed on previous trials will be beneficial when those items become targets on future trials. This is not the case in unrelated blocks. Here, residual activation will be left on words related to the target. Since those related words are not part of the set being produced, this residual activation will not affect production of future words. Therefore, semantic blocking may help those with segmental encoding deficits by adding activation that may help to overcome failures to activate the correct segments from the lexical node. This would lead to reduced training difficulty.

A related explanation of the reversed effect during training is that, as discussed in Chapter 1, both interference and facilitation exist within the word production system. For neurotypical adults, the facilitative effects of priming are quickly maxed out and masked by interference effects. However, in individuals with cognitive deficits, the balance of interference and facilitation may change so that the facilitative effects of priming outweigh the interference effects. Similarity priming may have benefits for performance similar to those for residual activation discussed above, providing additional activation that can aid selection and help overcome deficits like failures to activate.

Another possible explanation for why some participants were able to more accurately produce items in semantic vs. unrelated blocks during training is strategy use. Perhaps the appropriate distinction is not about spelling deficit but rather about cognitive control. For some reason, maybe successful individuals are able to use the fact that items are presented in related sets to their advantage. The present study did not include cognitive control or general learning and memory measurements, so more investigation would be needed to evaluate this possible explanation.

Segmental Blocking

The e-ILM predictions also address the effects of training in segmentally related blocks.

Segmental prediction 1: Training in segmental blocks leads to interference during acquisition relative to training in unrelated contexts. This prediction was investigated by comparing the improvement for items trained in semental blocks and unrelated blocks as participants practiced producing each item more times within a training session, much as the corresponding prediction about training in semantic blocks vs. unrelated contexts was evaluated. According to the prediction, participants should show smaller increases in accuracy during training for items in segmental blocks than for items in unrelated blocks. In direct contrast to this prediction, results of the study showed that items trained in segmental blocks had greater increases in accuracy during training than items trained in unrelated blocks for the group of participants as a whole. The same positive effect indicating facilitation as opposed to interference for training in segmental vs. unrelated blocks was observed for each of the seven participants when individual effects were examined. The prediction did not hold for this group of individuals with dysgraphia. Not only do the results of this study go against the prediction of the e-ILM, but they also go against the results of the previous study of neurotypical adults learning new words in which the predicted interference was seen.

Segmental prediction 3: Training in segmental blocks is detrimental for retention relative to training in unrelated contexts, assuming that reduced distinctiveness leads to worse long-term learning outcomes. This prediction was examined by comparing accuracy for the spellings of items trained in segmental blocks vs. unrelated contexts on recall assessments administered immediately post-training and at follow-up several weeks later. The prediction states that participants should show worse retention for items trained in segmental blocks than for items trained in unrelated contexts. The results were consistent with this prediction. Although there was still an increase in accuracy at follow-up relative to the pre-treatment assessment, there was less retention of items trained in segmental block than of items trained in unrelated contexts betweeen the immediately post-training assessment and the follow-up assessment. In looking at individual differences, only one participant showed the reversed effect of better retention for items trained in segmental blocks vs. unrelated contexts. Although there were null effects of segmental blocking on retention in the previous study with neurotypical learners, the results of the present study align with the prediction of the e-ILM: training in segmental blocks reduced retention relative to training in unrelated contexts.

Segmental prediction 4: Segmental blocking is not a desirable difficulty: increased training difficulty (interference) leads to reduced long-term retention due to reduced distinctiveness. This prediction was assessed in two ways. First, the pattern of results across participants was considered. According to the prediction, participants as a whole should show interference during training as well as a disadvantage in retention for items trained in segmental blocks as opposed to unrelated contexts. As described above, this was not the case. Although training in segmental blocks vs unrelated contexts led to reduced retention of those items as predicted, it also reduced training difficulty instead of increasing it as predicted. The effects on training and retention at the group level were in opposite directions, which is consistent with the pattern of a desirable difficulty. This contradicts the prediction of the e-ILM as well as the results of the previous study with neurotypical learners. However, there was not a corresponding effect in the other analysis evaluating this prediction, which examined the correlations between individual differences in segmental blocking effects on training and retention. At the individual level, there was not a consistent relationship between training and retention effects, somewhat weaking the claim that segmental blocking fits the profile of a desirable difficulty.

Discussion of segmental blocking. Overall, the results of this study offer only limited support for the e-ILM predictions regarding segmental blocking: the e-ILM does not seem to adequately describe the effects of training in segmental blocks vs. unrelated contexts for this group of individuals with dysgraphia. Training in segmental blocks as opposed to unrelated contexts led to increased improvement during training as opposed to the interference predicted through the application of the incremental learning model of word production to training of unknown words. While distinctiveness was not directly evaluated in this study, the failure to find interference during training (and the opposite finding of facilitation) suggests that the model's predictions about its reduction may not hold for this population. In line with the e-ILM prediction, training in segmental blocks as opposed to unrelated contexts did lead to reduced retention. However, this was not necessarily a confirmation of the model. This prediction makes the assumption that long-term learning outcomes are a result of distinctiveness changes, and it is not clear that distinctiveness was reduced as expected. The retention disadvantage may have a different cause. The final prediction of the e-ILM connects the other three predictions to the learning literature. It suggested that segmental blocking should not be a desirable difficulty: training in segmental blocks vs. unrelated contexts should lead to interference during training and reduced retention because it reduces distinctiveness. This prediction was not upheld, again perhaps because the expected effects on distinctiveness were not upheld. At the group level, segmental blocking fit the profile of a desirable difficulty, with opposite effects on training and retention. While this does contradict the prediction of the e-ILM, it is consistent with the alternative general hypothesis of desirable difficulties which claims that manipulations that have negative effects on training (i.e., increase training difficulty) have positive effects on long-term retention and vice versa.

It is important to remember that even though segmental blocking fit the profile of a desirable difficulty in that opposite effects on training and retention were observed, it was not beneficial for this group of individuals with dysgraphia. Segmental blocking reduced training difficulty relative to training in unrelated contexts both at the group level and for all seven participants at the individual level, clearly demonstrating that it did not increase difficulty. Furthermore, items that were trained in segmental blocks were at a disadvantage relative to items trained in unrelated contexts when long-term retention was tested both at the group level and for six of the seven participants at the individual level. That is, this manipulation neither increased difficulty nor had "desirable" effects on long-term learning. Why might individuals experience this pattern? Although this group of individuals had complex deficits that were not idea for investigating deficit-behavior relationships and further research is clearly needed, it is possible to speculate about why these participants might demonstrate these effects.

One reason that segmental blocking might lead to reduced training difficulty relative to training in unrelated contexts is that within segmental blocks there will be residual activation for the shared segments along with strengthening of connections between lexical nodes and the shared segments that may help overcome failures to activate the correct lexical nodes or segments. This would make it easier to produce those shared segments. For a person who initially cannot produce many or any of the letters, this facilitated production of the shared letters will lead to improved accuracy during training. Similarly, it could be that the facilitative effects of priming outweigh the interference due to segmental overlap within the damaged system, representing a different balance than in the neurotypical system.

Another possible reason for the reduced training difficulty for training of items in segmental vs. unrelated blocks is that participants may be able to use a conscious strategy of producing the shared letters if they realize all words in the block contain them. (This would be a reliable strategy here, unlike in the previous experiment with neurotypical participants in which overlap was distributed instead of predictable.) This might free cognitive resources to focus on the unshared letters, leading to producing them more accurately. When all trained words are mixed together, these advantages disappear. There will no longer be residual activation for shared segments, and the strategy of producing shared letters within a specific context will fail. Without the advantages, retention may be worse for these items than for items not trained in segmental blocks. This is especially true if segmental blocking truly is a desirable difficulty: if retention is increased to the extent that training difficulty is increased, then the reduced difficulty during training due to strategy use for these items presented in segmental blocks vs. unrelated contexts will lead to reduced retention.

Summary of Predictions and Evidence from This Study

The e-ILM predictions and the results of the study of individuals with dysgraphia relearning of spelling that bear on these predictions are listed in Table 13.

Table 13. Predictions of the e-ILM regarding the effects of training in semantic and segmental blocks, along with results of Study 2 that were used to evaluate these predictions.

Semantic Blocking

Prediction	Result	
1. Training in semantic blocks leads to interference during acquisition relative to training in unrelated contexts.	Not Supported: no significant differences in improvement within session for items trained in semantic vs. unrelated blocks for the group as a whole	✖
2. Training in semantic blocks increases the distinctiveness of representations by strengthening the connections between distinctive semantic features and lexical nodes while weakening the connections between shared semantic features and lexical nodes.	Not Directly Addressed	
3. Training in semantic blocks is beneficial for retention relative to training in unrelated contexts, assuming that increased distinctiveness leads to better long-term learning outcomes.	Not Supported: no significant differences in retention after training in semantic blocks vs. unrelated contexts for the group as a whole	✖
4. Semantic blocking is a desirable difficulty: increased training difficulty (interference) leads to increased long-term retention due to increased distinctiveness.	Supported: correlation of individual differences shows relationship between training difficulty and retention advantage for items trained in semantic blocks vs. unrelated contexts	✔

Segmental Blocking

Prediction	Result	
1. Training in segmental blocks leads to interference during acquisition relative to training in an unrelated context.	Not Supported: greater improvement within session for items trained in segmental blocks than in unrelated contexts for the group as a whole	✖
2. Training in segmental blocks reduces the distinctiveness of representations by strengthening the connections between lexical nodes and shared segments while weakening the connections between lexical nodes and distinctive segments.	Not Directly Addressed	
3. Training in segmental blocks is detrimental for retention relative to training in unrelated contexts, assuming that reduced distinctiveness leads to worse long-term learning outcomes.	Supported: reduced retention of items trained in segmental blocks as opposed to unrelated contexts for the group as a whole. However, no evidence that this is due to reduced distinctiveness	✔
4. Segmental blocking is not a desirable difficulty: increased training difficulty (interference) leads to reduced long-term retention due to reduced distinctiveness.	Not Supported: segmental blocking had opposite effects on training and retention for the group as a whole, consistent with the pattern of a desirable difficulty	✖

Overall, the outcomes of the analyses presented in this chapter provided evidence against the e-ILM predictions that resulted from directly extending the incremental learning model of word production to training of unknown words. Although most of the results of the previous study with neurotypical individuals provided support for the predictions of the model, it does not seem to apply across this group of individuals with dysgrapia. Neither semantic nor segmental blocking led to interference at the group level during training: improvement within training sessions did not differ for items trained in semantic blocks as opposed to unrelated contexts, and items trained in segmental blocks in fact showed more improvement within training sessions than items trained in unrelated contexts. Although distinctiveness was not directly assessed in the present study, the failure to find interference due to training in blocks suggests that distinctiveness may not be enhanced for items trained in semantic blocks and reduced for items trained in segmental blocks as predicted.

The other e-ILM predictions extend the predictions from training to long-term learning. Since the expected effects were not found for training in either semantic or segmental blocks as opposed to unrelated contexts, it is more difficult to evaluate these predictions. If blocking does not affect training as expected, it might not affect distinctiveness as expected, and therefore may not affect retention as expected. However, analysis did reveal important relationships between training and retention effects that speak to the learning literature more generally. Both semantic and segmental blocking fit the pattern of a desirable difficulty, with opposite effects on training and retention. This fits with the general hypothesis proposed in the learning literature: manipulations that increase training difficulty also lead to increased retention. For semantic blocking, this result also fits with the e-ILM predictions, which suggested that increased distinctiveness has opposite effects on training and retention. For segmental blocking, the result contradicts the e-ILM predictions, which suggested that reduced distinctiveness has negative effects on both training and retention.

In interpreting these results, it is also important to note that while both semantic and segmental blocking fit the pattern of desirable difficulties with opposite effects on training and retention, most participants experienced the detrimental side of this relationship. That is, blocking did not induce difficulty during training or have positive long-term outcomes for most individuals. Rather, training was facilitated and retention was reduced when items were practiced in blocks rather than unrelated contexts. Further work is needed to investigate why these effects were reversed for many of the participants and to better understand how different deficits relate to the effects of training in blocks vs. unrelated contexts.

Practical Implications of the Study

Beyond its theoretical implications for testing the e-ILM predictions, this study also has some practical implications for the treatment of dysgraphia. First, the study provides an example of a successful application of the Test-Study-Test protocol to dysgraphia treatment. The study did effectively train participants to produce the practiced items. Participants were more accurate in spelling the trained items both immediately after training and at follow-up several weeks later, although there was some loss of information between the post-training time points.

Second, this method of training had primarily item-specific effects: the group generalization analysis suggested that there was larger improvement over the course of the study for trained as compared to untrained items, and furthermore the group of untrained items did not improve from the beginning to the end of the study. While an ideal treatment would lead to generalization to untrained items, an item-specific treatment like this one can still be considered successful and have a positive impact on the lives of the participants as they use the trained words in day-to-day life.

Third, the opposite directions of the effects of blocking for training and retention point out that it is unwise to generalize from the effects during training to the effects during recall: a variable that positively affects training will not necessarily positively affect retention, so both need to be considered in the evaluation of treatment efficacy.

Fourth, practitioners should be cautious about training items in related blocks. It is clear that neither semantic blocking nor segmental blocking should be applied to all participants. While both types of blocking fit the profile of a desirable difficulty, training in semantic or segmental blocks relative to unrelated contexts had a negative long-term impact for most participants. Just because a variable fits the profile of a desirable difficulty and has positive long-term effects in neurotypical participants does not mean it should be universally applied in treatment of individuals with neurological damage. The specific deficits of the participant need to be taken into account. More research is needed to determine the deficit profiles of those who are most likely to have positive as opposed to negative long-long term effects of each type of blocking.

CHAPTER 6: SUMMARY & CONCLUSIONS

Synopsis

Background Work Supports Incremental Learning Model of Word Production

The work presented in this dissertation enhances our understanding of the language production and learning systems. Background experiments investigating the nature of the intact production system showed that interference arises not only as a result of semantic similarity, but also as a result of segmental similarity when it is not predictably limited to the first segment. Furthermore, consistent effects are found across written and spoken production. I argued that incremental learning models like those presented by Howard and colleagues (2006) and by Oppenheim and colleagues (2010) to explain lexical selection can be extended to segmental encoding in order to account for these effects. Moreover, I suggested that the same principles underlie both stages of production across the written and spoken modalities.

Extending the Incremental Learning Model to Word Learning

In this dissertation, I extended this work from production of known words to learning of new words. When a person learns a new word, it must be integrated into the production system so that it can be used. Here, I have suggested that this process relies on the same learning mechanisms that are involved in the production of known words, applying an incremental learning model to learning words in addition to producing them. In extending the model, I also related theories of long-term learning and retention regarding the effects of difficulty during learning on retention and the need to create distinctive representations to theories of word production. This novel application of the model led to clear, testable predictions about the effects of learning new words in the context of other similar words that are similar, which had not been studied prior to the work presented here. The studies presented in the dissertation are examples of confirmatory science, investigating whether the e-ILM predictions correctly describe the effects of semantic and segmental blocking on the learning of new words by neurotypical individuals and on the relearning of spellings by individuals with dysgraphia. Below, I present the predictions followed by the results of the studies that evaluated them. This information also summarized in Table 14.

Predictions regarding semantic blocking. Past studies have suggested that semantic similarity may be a desirable difficulty that leads to interference during learning but has long-term benefits for the outcomes of learning. Extending the incremental learning model proposed by Oppenheim and colleagues (2010) to this situation offered a potential explanation for this desirable difficulty: increased distinctiveness. There is interference during learning of items in semantic blocks relative to learning of items in unrelated contexts, just as there is for production of known words from the same category, as a result of repeated weakening of connections between shared features and lexical nodes. Over time, however, semantic similarity has positive effects on the representations of trained words because distinctiveness is enhanced as a result of repeated weakening of connections between distinctive features and lexical nodes. That is, over the course of learning, distinctive features that can be used to differentiate between the related items are strengthened while the shared features that apply to all are weakened. Distinctiveness is not systematically modified in the same way when items are presented in unrelated contexts. Further extending the predictions to effects on long-term learning, people should show better retention for words that are learned in semantic blocks than for words that are learned in unrelated contexts as a result of the enhanced distinctiveness for items presented in semantic blocks vs. unrelated contexts. Relating these predictions to the learning literature, in this situation blocking is a desirable difficulty. Because of enhanced distinctiveness, training in semantic blocks has opposite effects on training and retention as compared to training in segmental blocks.

Predictions regarding segmental blocking. Past studies offer a less clear picture regarding segmental similarity, although the balance of evidence suggests that it may not be a desirable difficulty. The extension of the incremental learning model again offers distinctiveness as a possible explanation for the pattern. Presenting new words in segmental blocks leads to interference during learning relative to presentation in unrelated contexts, just as it does for production of segmentally related known words. This is a result of weakening connections between lexical nodes and distinctive segments while strengthening connections between lexical nodes and shared segments. Instead of enhancing distinctiveness, this type of similarity reduces it: over the course of learning, distinctive segments that can be used to differentiate between related items are weakened while shared segments that are part of many or all of the related items are strengthened. Distinctiveness is not modified in the same way when items are presented in unrelated contexts. Again extending the predictions to long-term learning, the resulting reduction in distinctiveness for items learned in segmental blocks as opposed to unrelated contexts leads to worse retention of those items. Therefore, in this situation blocking is not a desirable difficulty, in contrast to the general hypothesis of desirable difficulty that suggests that any manipulation that increases training difficulty should also improve long-term learning outcomes. Rather, training in segmental blocks as opposed to unrelated contexts has detrimental effects on both training and retention because it leads to reduced distinctiveness.

Evaluating the Predictions in Neurotypical Learning of New Words

I first tested these predictions in Chapter 4, which presented a study regarding neurotypical learning of new words. In this experiment, healthy young adults were trained on twenty-four novel items that were presented in blocks that were semantically related, segmentally related, and unrelated. Separate groups of individuals participated in written and spoken versions of the experiment, which resulted in similar patterns of effects, suggesting similar underlying mechanisms. I not only looked at the speed and accuracy with which new words are produced after learning in order to investigate whether learning in similar contexts presents desirable difficulties, but I also directly examined whether potential differences between effects of semantic and segmental similarity result from changes in the distinctiveness of representations by probing associations between the new items and distinctive vs. shared semantic features and segments.

Semantic blocking. In line with the predictions of the e-ILM, training in semantic blocks as opposed to unrelated contexts increased training difficulty and distinctiveness of the representations. Extensions of the predictions to long-term learning were less strongly supported. Semantic blocking led to improved retention to some extent in the spoken experiment in that participants responded more quickly to words trained in semantic blocks vs. unrelated contexts when recall was tested several weeks after training, although significant effects on retention were not observed in the written experiment or in analyses evaluating retention across training sessions. Individual differences in retention and training effects were negatively correlated in the written experiment such that semantic blocking increased retention of the items to the extent that it also increased training difficulty. Together, the overall pattern of results in the spoken experiment and the correlations in the written experiment suggest that semantic blocking is in fact a desirable difficulty. The e-ILM predictions accurately describe the consequences of training in semantic blocks as opposed to unrelated contexts for this group of neurotypical participants, although the predictions resulting from directly applying the incremental learning model of word production to training of new words are more strongly supported than those extending it to long-term learning.

Segmental blocking. Results regarding segmental blocking were less straightforward. In line with the e-ILM predictions, segmental blocking as compared to training in unrelated contexts increased training difficulty and reduced distinctiveness. However, the predictions relating training to long-term learning were not as clearly supported. In contrast to the predictions, there was no evidence of reduced long-term retention for items trained in segmental blocks vs. unrelated contexts. The e-ILM prediction that training in segmental blocks relative to unrelated contexts would both increase difficulty during training and reduce retention was not upheld in either the overall pattern across participants or in the analysis of individual differences. The failure to find a negative impact of training in segmental blocks on retention relative to training in unrelated context may be because the effect whereby training in segmental blocks reduced distinctiveness was relatively small, making it more difficult to find effects on retention. A different manipulation of segmental similarity might increase the size of the effect. Another possibility is that the effects of segmental similarity on retention might require a longer time period to emerge; future studies should track long-term recall over an increased retention period. Additionally, the present study may suffer from low power: a larger number of items and/or participants may be needed to examine retention effects. Alternatively, it could be that the predictions of the incremental learning model are correct in predicting that training in segmental blocks increases training difficulty and reduces distinctiveness, but that the assumption that reduced distinctiveness leads to worse long-term learning outcomes is incorrect. Neurotypical individuals may be well equipped to deal with the limited distinctiveness caused by segmental similarity and therefore may still be able to learn new words that are segmentally similar. This is because language relies on relatively small closed classes of segments that repeatedly appear across words in the language in various combinations. If people could not make the necessary distinctions between words with similar segments, their production would be extremely limited. Since individuals have a great deal of practice making the necessary distinctions, the impact of training in segmentally similar blocks may be less negative than predicted. In fact, there may be advantages of directly contrasting similar words within a set to help reinforce the differences between items, especially when feedback is given. Some language-external mechanism may therefore allow people to overcome the interference induced by segmental similarity.

Overall, the results of the first study were generally in line with the more direct e-ILM predictions regarding the effects of blocking on training difficulty and distinctiveness. The model of training does seem to apply for this group of neurotypical learners. However, results addressing the extensions of the predictions to long-term learning outcomes provided less support, chiefly because there were only weak effects of blocking on retention. Further research is needed to investigate why the expected retention effects were not observed, considering the possibilities of imperfect methodology, incorrect assumptions connecting distinctiveness and long-term learning outcomes, and language-external mechanisms used to overcome interference. Better understanding of potential retention effects is also needed to more completely evaluate the predictions regarding desirable difficulties. While the present study suggests that semantic blocking is likely to fit the profile of a desirable difficulty but segmental blocking may not, these results are not conclusive and would be clarified by more complete information about retention effects due to training in blocks vs. unrelated contexts.

Evaluating the Predictions in Treatment of Individuals with Dysgraphia

In Chapter 5, I looked at a further practical application of this research, investigating the effects of treating items in related contexts on the relearning of spellings by individuals with dysgraphia. This provided an opportunity to test the predictions of the incremental learning model and to see if the same general effects are found in these individuals with cognitive deficits as are found in neurotypical individuals. Seven individuals with complex spelling deficits participated in a treatment study in which they received treatment for thirty-six items that were presented in semantically related, segmentally related, or unrelated blocks.

Semantic blocking. In this treatment study, the first prediction of the e-ILM was regarding direct application of the incremental learning model of word production to training was not upheld: for the group as a whole, there was not interference during training when items were practiced in semantic blocks as opposed to unrelated contexts. Although distinctiveness was not directly tested, this result suggested that distinctiveness was not enhanced as predicted for this group of participants. Given that distinctiveness was likely not increased by this training, it is perhaps unsurprising that the prediction addressing long-term learning outcomes was not supported either: there was no overall effect on retention of training in semantic blocks vs. unrelated contexts. However, these results did provide the opportunity to examine the relationship between training and retention effects. Semantic blocking was shown to fit the profile of a desirable difficulty. Individual differences in effects on training and retention were correlated so that the individuals who experienced the greatest difficulty during training when items were presented in semantic blocks as opposed to unrelated blocks were the same individuals who showed the greatest retention of the items trained in semantic blocks as opposed to unrelated blocks. This result is consistent with the e-ILM's predictions and account of the underlying mechanisms. For some individuals, training in semantic vs. unrelated blocks did increase training difficulty and correspondingly increase retention, potentially because this type of training enhanced distinctiveness as proposed by the model. However, for others with damaged spelling systems, this was not the case, which may have masked overall effects. Note that these results are also consistent with the general hypothesis of desirable difficulties whereby manipulations that increase training difficulty also lead to improved long-term learning outcomes.

Segmental blocking. As with semantic blocking, the first e-ILM prediction regarding segmental blocking was not supported by the results of this study. In contrast to the interference predicted during training of items in segmental blocks vs. unrelated contexts, facilitation was found. That is, for all individuals in this study and for the group as a whole, training in segmental blocks reduced training difficulty as compared to training in unrelated contexts. While distinctiveness was not probed in this study, this result suggests that training in segmental blocks may not reduce distinctiveness for this group of participants. In line with the e-ILM prediction that extended the model to long-term learning outcomes, reduced retention was observed for words trained in segmental vs. unrealted contexts for 6/7 participants at the individual level and for the group as a whole. However, it is not clear that this reduced retention is truly due to reduced distinctiveness. In relating the results of this study to the learning literature, segmental blocking fit the profile of a desirable difficulty: it had opposite effects on training and retention. This contradicts the e-ILM prediction, but it is consistent with the general hypothesis of desirable difficulties. This principle of learning seems to apply to segmental blocking as well as semantic, at least in this specific group of participants with spelling deficits who did not display the training effects predicted by the e-ILM. More research is needed to determine how specific deficits may relate to the observed effects, potentially explaining why the effects observed in neurotypical participants learning new words were not consistent with the effects observed for this group of individuals with dysgraphia relearning spellings.

Summary.   The e-ILM predictions and the results of the two studies that evaluated them are presented in Table 14.

Table 14. Predictions of the e-ILM regarding the effects of training in semantic and segmental blocks, along with results of both studies that were used to evaluate these predictions.

Semantic Blocking

Prediction	**Results of Study 1: Learning of New Words in Neurotypical Adults**	**Results of Study 2: Treatment of Spelling in Individuals with Dysgraphia**
1. Training in semantic blocks leads to interference during acquisition relative to training in unrelated contexts.	Supported: greater improvement within session for items trained in unrelated contexts than in semantic blocks	Not Supported: no significant differences in improvement within session for items trained in semantic vs. unrelated blocks for the group as a whole
2. Training in semantic blocks increases the distinctiveness of representations by strengthening the connections between distinctive semantic features and lexical nodes while weakening the connections between shared semantic features and lexical nodes.	Supported: advantage for verification of distinctive features vs. shared features of items trained in semantic blocks on semantic probe tasks	Not Directly Addressed
3. Training in semantic blocks is beneficial for retention relative to training in unrelated contexts, assuming that increased distinctiveness leads to better long-term learning outcomes.	Partially Supported: faster responses to items trained in semantic blocks vs. unrelated contexts at follow-up in the spoken experiment	Not Supported: no significant differences in retention after training in semantic blocks vs. unrelated contexts for the group as a whole

4. Semantic blocking is a desirable difficulty: increased training difficulty (interference) leads to increased long-term retention due to increased distinctiveness.	Partially Supported: pattern of results across subjects of interference during training and potential retention advantage at follow-up for items trained in semantic blocks vs. unrelated contexts in spoken experiment; correlation of individual differences shows relationship between training difficulty and retention advantage over sessions for items trained in semantic blocks vs. unrelated contexts in written experiment	Supported: correlation of individual differences shows relationship between training difficulty and retention advantage for items trained in semantic blocks vs. unrelated contexts

Segmental Blocking		
Prediction	**Results of Study 1: Learning of New Words in Neurotypical Adults**	**Results of Study 2: Treatment of Spelling in Individuals with Dysgraphia**
1. Training in segmental blocks leads to interference during acquisition relative to training in an unrelated context.	Supported: greater improvement within session for items trained in unrelated contexts than in segmental blocks	Not Supported: greater improvement within session for items trained in segmental blocks than in unrelated contexts for the group as a whole
2. Training in segmental blocks reduces the distinctiveness of representations by strengthening the connections between lexical nodes and shared segments while weakening the connections between lexical nodes and distinctive segments.	Supported: disadvantage for verification of distinctive features vs. shared features of items trained in segmental blocks on segmental probe tasks	Not Directly Addressed
3. Training in segmental blocks is detrimental for retention relative to training in unrelated contexts, assuming that reduced distinctiveness leads to worse long-term learning outcomes.	Not Supported: no significant differences in retention after training in segmental blocks vs. unrelated contexts	Supported: reduced retention of items trained in segmental blocks as opposed to unrelated contexts for the group as a whole. However, no evidence that this is due to reduced distinctiveness
4. Segmental blocking is not a desirable difficulty: increased training difficulty (interference) leads to reduced long-term retention due to reduced distinctiveness.	Not Supported: pattern of results across subjects of interference during training but no effect on retention of training in segmental blocks vs. unrelated context; no significant correlation between training and retention effects when individual differences considered.	Not Supported: segmental blocking had opposite effects on training and retention for the group as a whole, consistent with the pattern of a desirable difficulty

Implications of the Study

Together, the two studies presented in this dissertation have theoretical implications for understanding the system used to learn new words, as well as practical implications for the training of words in rehabilitation and education contexts.

First, the predictions of the e-ILM that directly applied an incremental learning model of production to the training of unknown words were supported by the results of the study of neurotypical adults learning new words. Training in both semantic and segmental blocks lead to interference relative to training in unrelated blocks. While both types of blocking increased training difficulty, they had different impacts on the distinctiveness of the representations of the words being learned. Training in semantic blocks increased distinctiveness while training in segmental blocks reduced distinctiveness as shown by the results of the distinctiveness probe tasks. Based on these results, the e-ILM accurately described the effects of training new words in semantic and segmental blocks, suggesting that this is a valid model that may offer insight into the process of learning new words in the neurotypical system.

However, the same effects were not observed in the second study, which examined individuals with dysgraphia relearning the spellings of previously known words. In contrast to the predictions and results of the previous study with neurotypical individuals, no interference was observed during training for items practiced in semantic blocks vs. unrelated contexts, and facilitation was observed during training for items practiced in segmental blocks vs. unrelated contexts. Distinctiveness was not directly assessed. The deficits of the individuals with dysgraphia disrupted the system in a way that changed the effects of blocking, contradicting the predictions of the e-ILM.

The predictions of the e-ILM were also extended beyond training effects to address long-term learning outcomes. For these predictions, the situation is different. This time, it was the results of the study with neurotypical participants that did not align with the predictions. Although, as discussed above, training in semantic blocks increased distinctiveness while training in segmental blocks decreased distinctiveness, there were not strong effects on retention of the items. That is, the predicted retention advantage for items trained in semantic blocks vs. unrelated contexts was only weakly supported, and the predicted retention disadvantage for items trained in segmental blocks vs. unrelated contexts was not observed at all. While the failure to find consistent retention effects may have been due to the specific methodological choices made in the present study or due to language-external factors in the learning system, it also calls into question the assumption that increased distinctiveness leads to better long-term learning outcomes. Future research is needed to examine these possibilities and more fully characterize the effects of training in blocks as opposed to unrelated contexts on retention.

On the other hand, the results of the study with individuals with dysgraphia provided support for a general learning principle relating training and retention effects. Both semantic and segmental blocking fit the profile of desirable difficulties, meaning that they had opposite effects on training and retention. To the extent that blocking increased training difficulty relative to training in unrelated contexts, it also increased retention of the items trained in blocks. This finding was expected for semantic blocking: it aligned with the predictions of the e-ILM. There were also some indications consistent with semantic blocking acting as a desirable difficulty in the study with neurotypical participants when examining the overall pattern of results in the spoken experiment and the relationship between individual differences in training and retention effects. However, the e-ILM predicted that segmental blocking would not be a desirable difficulty, but rather would increase training difficulty and reduce retention relative to training in unrelated contexts. More investigation is needed to determine whether the observed pattern truly reflects a general learning principle whereby increased difficulty during training leads to increased retention, or whether it was a result of the specific deficits of the individuals in this study. That is, further research is needed to evaluate whether segmental blocking acts as a desirable difficulty for neurotypical individuals.

The contrasting results between the studies with neurotypical adults and individuals with dysgraphia raise potentially important questions about learning and relearning. Do the differences arise because learning processes are fundamentally changed after brain damage? That is, do the same underlying mechanisms apply for both neurotypical adults and those with cognitive deficits, or does damage to the word production system change the processes involved in learning words? Are the processes of learning new words and relearning previously known words different in some way? These questions could be addressed in future work by teaching individuals with cognitive deficits new words that they would not have known premorbidly (like the stimuli used in the neurotypical study). Performance on this task could be directly compared to performance on a relearning task in which previously known words are practiced in order to directly investigate differences in learning and relearning. It is also possible that learning and relearning are similar, and that learning processes are not fundamentally changed by damage. If this is the case, the contrasting results may be due to the specific conditions of the studies presented here. As mentioned above, further investigation of the effects of blocking on long-term learning in neurotypical individuals is required, as is examination of the relationships between specific cognitive deficits and behavioral performance.

There are also practical implications of the findings of these studies for rehabilitation. First, the results presented here show that it is important to consider different scales of learning when designing treatments. Manipulations that increase difficulty during treatment may initially seem detrimental, but they may lead to positive long-term outcomes and vice versa. Both short-term and long-term effects should be evaluated in determining the efficacy of a treatment protocol.

Second, applying the same treatment to all individuals, even if the treatment manipulation is beneficial for neurotypical individuals, may not result in positive effects for all. While semantic and segmental blocking both fit the pattern of desirable difficulties in the second study, blocking was not uniformly advantageous for individuals with dysgraphia as they relearned the spellings of previously known words. The cognitive profiles of the individuals must be considered when treatment is applied. The long-term benefits and disadvantages of blocking may depend on deficit characterization. Further research with individuals with more circumscribed deficits is needed to more fully characterize who benefits and who does not. The present study only evaluated spelling deficits; future research should also look at more general learning, memory, and control impairments to see how blocking interacts with them to affect word relearning, as well as investigating effects of blocking on spoken treatment for individuals with spoken language deficits. Although the similar patterns of results for the spoken and written modalities in neurotypical learning suggest that there are similar underlying mechanisms across the modalities, damage may affect them differently.

Additional research is also needed to determine the optimal amount of difficulty to introduce during training. Too little difficulty may have negative effects on long-term retention, but so may too much difficulty if the individual does not have the necessary background and skills to deal with that difficulty.

The work presented here may also have implications for education. Although future research is needed to determine if effects are the same for younger children whose language systems are still developing, the present evidence suggests that training in semantic or segmental blocks, at the very least, is not detrimental for neurotypical adults. This suggests that segmental and semantic similarity do not necessarily need to be avoided in vocabulary learning. Furthermore, results from the treatment study of dysgraphia suggest that both types of blocking may be desirable difficulties. This suggests that teaching new vocabulary in semantic or segmental blocks may be beneficial for long-term learning, even if it increases immediate difficulty during training. Future studies should be undertaken to directly investigate the effects of blocking in an educational context.

Overall, the work presented here enhances the understanding of the mechanisms involved in the production and learning of new words. Although further research is needed to clarify effects of semantic and segmental blocking on long-term retention, the e-ILM does seem to describe the effects of blocking during training accurately. The work presented here also suggests that the general principle of desirable difficulties applies to the learning of words. The contrasting results observed in the studies with neurotypical adults and individuals with dysgraphia raise questions about potential differences in learning and relearning that also merit future investigation.

APPENDICES

Appendix A: Extending the Howard et al. (2006) Account to Segmental Encoding

Consider applying the logic of Howard and colleague's (2006) model to segmental encoding, which involves the lexical and segmental levels of processing. The same principles of competition, priming, and shared activation apply as in the account of lexical selection. This model involves the same lexical level that is involved in lexical selection: therefore, lateral inhibition between lexical nodes proportionate to each item's activation remains (competition)[24]. Additionally, when a target is produced, the connections between the selected segments and lexical nodes that contributed to their activation are strengthened (priming). Because there is feedback between the lexical and segmental levels (e.g., Dell et al., 2014; Rapp & Goldrick, 2000), other lexical nodes that share segments are likely to be active as a target is being processed (shared activation). Figure A1 depicts an example of this model.

[24] Lateral inhibition between segments competing for the same position in the output may also be present, although it is not critical to the predictions stated here, but rather would augment the predicted interference.

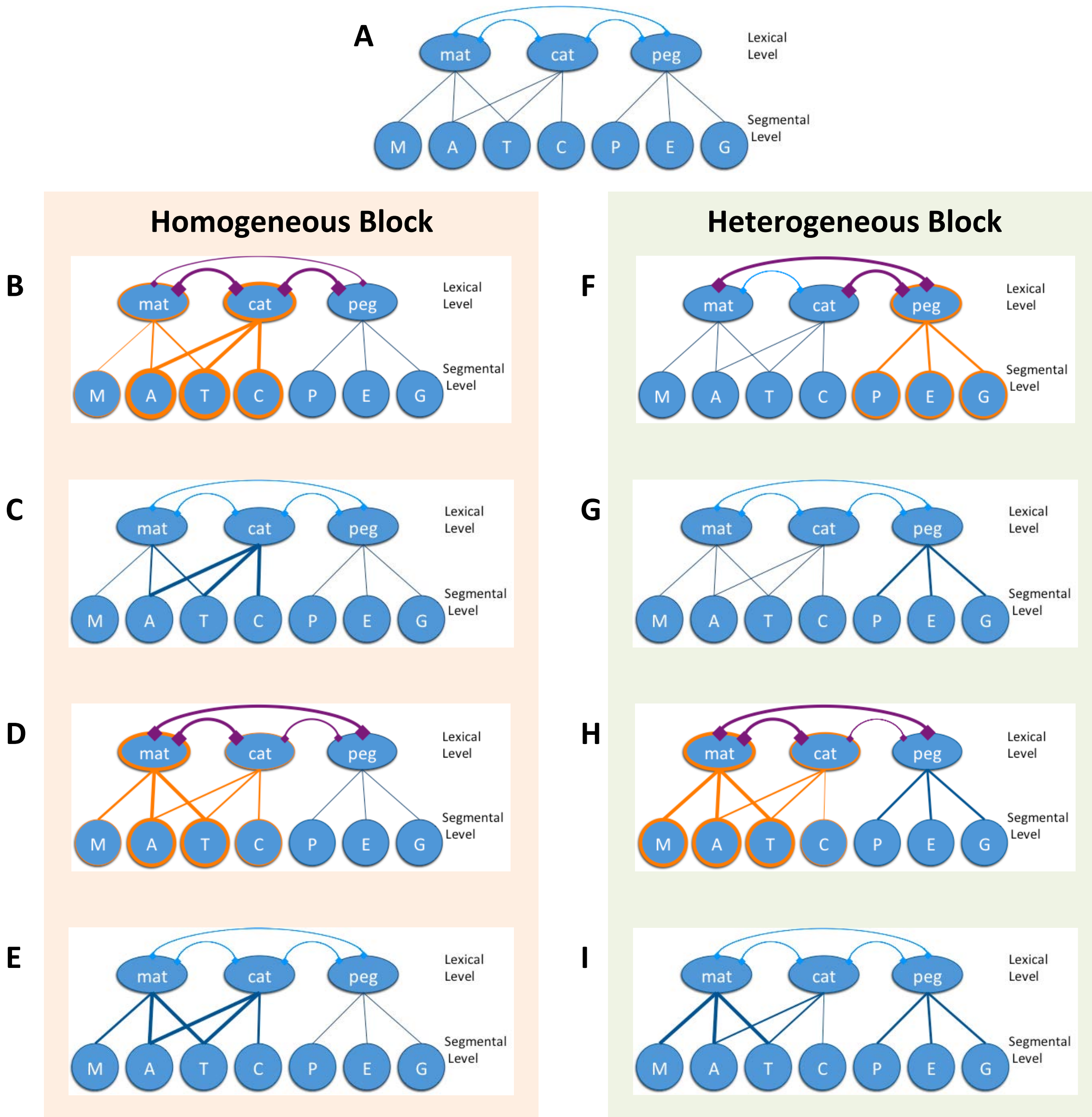


Figure A1. Example implementation of an extended version of the Howard and colleagues (2010) model that applies to segmental encoding.

In panel A, the initial state of the system is shown: there are three lexical nodes, with connections to seven segments. Two of the lexical nodes, *mat* and *cat*, are segmentally related, each sharing two of their three segments, while the third, *peg*, is not. Panels B-E, in the purple box, depict naming in a segmentally homogeneous block. Here, the segmentally related targets *cat* and then *mat* are named. Panels F-I, in the green box, depict naming in a segmentally heterogeneous block. Here, the segmentally unrelated targets *peg* and then *mat* are named. Orange lines depict activation during each trial. Purple lines between lexical nodes depict lateral inhibition. Blue lines depict the connections between levels. The weight of these lines is used to depict strengthening, with thicker, solid lines showing strengthened connections. Refer to the text for a detailed description of the example depicted in this figure.

Panel A the initial state of the system: there are three lexical nodes, with connections to seven segments. Two of the lexical nodes, *mat* and *cat*, are segmentally related, each sharing two of their three segments, while the third, *peg*, is not segmentally related to the others.

Panels B-E illustrate what happens in a segmentally homogeneous block where *mat* and *cat* are named.

In panel B, the lexical node *cat* is activated. This leads to activation of *cat*'s segments. Through feedback, these segments send activation to the lexical node *mat* that shares many of the same segments. *Mat* in turn sends activation to its segments, including M that is not part of the target *cat*. There is, however, lateral inhibition between the simultaneously active lexical nodes, so *cat* will inhibit *mat* and vice versa. The segments of *cat* receive more activation than those of *mat* since the unshared segment in *mat* only receives activation from feedback and because *cat* inhibits *mat* more strongly than *mat* inhibits *cat* since *mat* is less active than *cat.* The segments of *cat* can be selected since they are more active than all other competitors.

Panel C shows that when the correct target *cat* is produced, the connections that contributed to activation of its segments are strengthened. These strengthened connections include the ones between the target segments A and T and the non-target lexical node *mat* because all connections that contribute to activation of the correct target segments are strengthened.

Panel D shows what happens when the target of the next trial, *mat*, is segmentally related to the target of the first trial, *cat*. Activating the segments of *mat* also leads to the activation of *cat* through feedback. Because the connections between the lexical node *cat* and the shared segments were strengthened on the previous trial, it is a strong competitor: *cat* has greater activation and it inhibits *mat* more strongly than it would have had it not been the previous response. Therefore, there is interference: it takes longer to select *mat* after the related *cat* because *cat* is active and inhibiting it more than it would had *cat* not been the previously produced target.

Panel E shows that after production of the second target mat, there is again strengthening of connections between the selected segments and the lexical nodes that contributed to their activation.

Contrast panels B-E with panels F-I, which illustrate what happens in a segmentally heterogeneous block.

Panel F shows that when the lexical node of the first target, *peg*, is activated, it sends activation to its segments. Because there are no items in this example that share segments with *peg*, there is no feedback that leads to activation of non-target lexical nodes or segments. *Peg* is more active than other competitors and its segments are selected and produced.

Panel G shows that when the correct target *peg* is produced, the connections that contributed to activation of its segments are strengthened. In this example, there were no connections that contributed to activation of non-target segments.

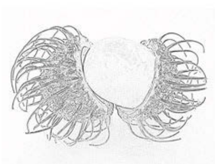
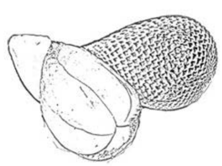
Panel H shows what happens when the target of the next trial, *mat*, is segmentally unrelated to the target of the first trial, *peg*. Here, activation of *mat* is unaffected by the previous trial. Since *mat* and *peg* do not share segments, the strengthened connections between *peg*'s lexical node and segments are not involved. In contrast to the situation presented in panel D, there is not interference between these items.

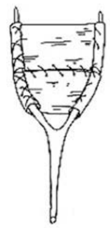
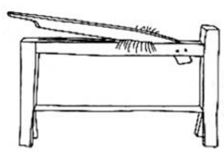
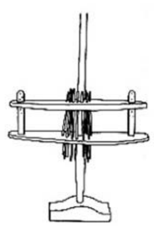
Panel I shows that after production of the second target *mat*, there is strengthening of connections between segments and the lexical nodes that contributed to their activation. This strengthening does not interact with the previous strengthening of *peg*'s connections.

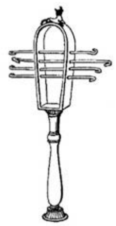
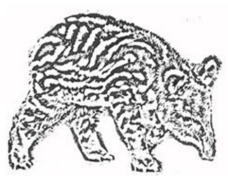
According to this account, a second target that is segmentally related to the first target is at a disadvantage compared to one that is unrelated to the first target because the related word is a stronger competitor (e.g., *cat* will be more active and inhibit *mat* more than *peg* will be active and inhibit *mat*): production will be slower and/or more error prone when the second word is related to the first than when the second word is unrelated to the first. This model predicts interference will be observed as increased response time and/or reduced accuracy when words are named in the context of segmentally related vs. unrelated items.

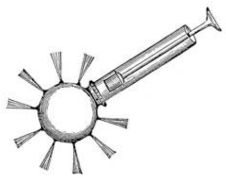
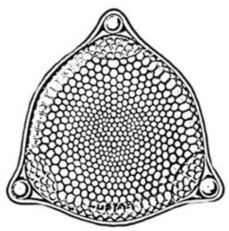
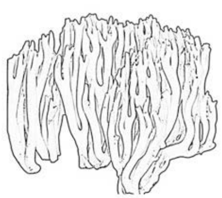
.

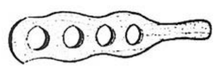
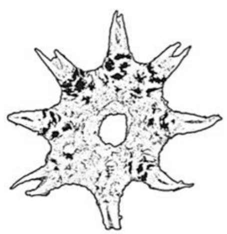
Appendix B: Stimuli Used in Study 1

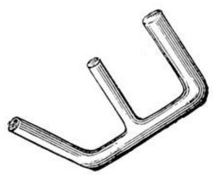
Group: Semantic1	Picture	Written experiment spelling	Written experiment pronunciation	Spoken experiment spelling	Spoken experiment pronunciation	Features
		chys	JIs	crube	krub	Green Tastes sweet Used for food Consumed cooked
		fulb	fVlb	dwipe	dw2p	Red Tastes sour Used for food Consumed raw
		vork	v9k	haint	h1nt	Orange Tastes sweet Used for medicine Consumed raw
		jing	_IN	smuth	smVT	Black Tastes sweet Used for intoxication Consumed raw

Group: Sematic 2	Picture	Written experiment spelling	Written experiment pronunciation	Spoken experiment spelling	Spoken experiment pronunciation	Features
		frop	frQp	drage	dr1_	3 feet long Wooden Removes seeds Used with wheat
		wung	wVN	glybe	gl2b	1 foot long Metal Removes seeds Used with flax
		jalt	_{lt	skeph	skEf	4 feet long Wooden Separates straw Used with flax
		byme	b2m	shump	SVmp	2 feet long Wooden Spins fibers into yarn Used with flax

Group: Segmental 1	Picture	Written experiment spelling	Written experiment pronunciation	Spoken experiment spelling	Spoken experiment pronunciation	Features
		fusp	fVsp	bleck	blEk	Bronze 18 inches wide Percussion instrument Used in worship
		fing	fIN	bline	bl2n	6 inches long Silver Cuts paper Used by women
		lisk	lIsk	brudd	brVd	Weighs 50 pounds Shades of brown Herbivore Swims
		rish	rIS	gluth	glVT	15 feet tall Bitter fruit Used for timber Used for crafts

Group: Segmental 2	Picture	Written experiment spelling	Written experiment pronunciation	Spoken experiment spelling	Spoken experiment pronunciation	Features
		sove	s5v	swush	swVS	Blue Rubber Bends Toy
		samb	s{m	pluff	plVf	Glass 8 inches long Uses pressure Disperses liquid
		jace	_1s	slaim	sl1m	Microscopic Yellow Found in fresh water Cannot move
		yate	j1t	sloot	slut	Magenta 7 inches tall Found in grassland Consumes dead matter

Group: Unrelated 1	Picture	Written experiment spelling	Written experiment pronunciation	Spoken experiment spelling	Spoken experiment pronunciation	Features
		virt	v3t	twish	twIS	Gold Weighs 5 pounds Container Used in 1800s
		musp	mVsp	glafe	gl1f	Tan Weighs 1/2 pound Used by children Used in 2 player game
		chen	JEn	scoon	skun	Gray Weighs 10 pounds Found in deep sea Eats crustaceans
		galk	g{lk	themp	TEmp	Pink Smells of dung Found in tropics Attracts flies

Group: Unrelated 2	Picture	Written experiment spelling	Written experiment pronunciation	Spoken experiment spelling	Spoken experiment pronunciation	Features
		fomb	fQm	grive	gr2v	High pitched sound Ceramic Used for dancing music Held in both hands
		cive	s2v	dwesh	dwES	Plastic Rigid Attaches to tubing Used in laboratory
		kyth	kIT	bloof	bluf	Purple Very small Causes disease Used in making compost
		jasp	_{sp	thunt	TVnt	Pleasant scent White Rapid growth Weed

Appendix C: Psycholinguistic Characteristics of Pseudoword Stimuli Used in Study 1

Table C1. Psycholinguistic characteristics of pseudoword stimuli used in the written experiment. Values are from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002).

group	spelling	pronunciation	letters	phonemes	orthographic neighbors	summed frequency of orthographic neighbors	body neighbors	summed frequency of body neighbors
Semantic1	chys	JIs	4	3	2	1	1	1
Semantic1	fulb	fVlb	4	4	2	267	1	6
Semantic1	vork	v9k	4	3	5	893	6	894
Semantic1	jing	_IN	4	3	8	211	18	946
	group mean		4.00	3.25	4.25	343.00	6.50	461.75
Semantic2	frop	frQp	4	4	6	4068	23	602
Semantic2	wung	wVN	4	3	7	115	14	154
Semantic2	jalt	_{lt	4	4	4	64	4	64
Semantic2	byme	b2m	4	3	0	0	2	5
	group mean		4.00	3.50	4.25	1061.75	10.75	206.25
Segmental1	fusp	fVsp	4	4	3	22	1	0
Segmental1	fing	fIN	4	3	14	801	18	946
Segmental1	lisk	lIsk	4	4	5	167	4	76
Segmental1	rish	rIS	4	3	8	612	5	312
	group mean		4.00	3.50	7.50	400.50	7.00	333.50
Segmental2	sove	s5v	4	3	13	2368	19	683
Segmental2	samb	s{m	4	3	3	629	2	19
Segmental2	jace	_1s	4	3	10	649	12	1251
Segmental2	yate	j1t	4	3	9	531	17	905
	group mean		4.00	3.00	8.75	1044.25	12.50	714.50
Unrelated1	virt	v3t	4	3	2	19	6	85
Unrelated1	musp	mVsp	4	4	6	866	1	0
Unrelated1	chen	JEn	4	3	5	4147	18	5054
Unrelated1	galk	g{lk	4	4	7	357	5	362
	group mean		4.00	3.50	5.00	1347.25	7.50	1375.25

Unrelated2	fomb	fQm	4	3	4	55	4	55
Unrelated2	cive	s2v	4	3	11	1170	13	1018
Unrelated2	kyth	kIT	4	3	1	19	1	19
Unrelated2	jasp	_{sp	4	4	4	8	6	31
	group mean		4.00	3.25	5.00	313.00	6.00	280.75

group	spelling	pronunciation	body friends	summed frequency of body friends	body enemies	summed frequency of body enemies	onset neighbors	summed frequency of onset neighbors	phonological neighbors	summed frequency of phonological neighbors
Semantic1	chys	JIs	0	0	1	1	144	2420	14	9495
Semantic1	fulb	fVlb	1	6	0	0	447	24349	1	5
Semantic1	vork	v9k	5	143	1	751	77	873	12	702
Semantic1	jing	_IN	18	946	0	0	104	2545	17	2363
	group mean		6.00	273.75	0.50	188.00	193.00	7546.75	11.00	3141.25
Semantic2	frop	frQp	23	602	0	0	70	5428	8	3890
Semantic2	wung	wVN	14	154	0	0	361	52682	13	6935
Semantic2	jalt	_{lt	1	2	3	62	104	2545	2	2
Semantic2	byme	b2m	2	5	0	0	597	27918	21	2445
	group mean		10.00	190.75	0.75	15.50	283.00	22143.25	11.00	3318.00
Segmental1	fusp	fVsp	1	0	0	0	447	24349	2	0
Segmental1	fing	fIN	18	946	0	0	447	24349	20	2521
Segmental1	lisk	lIsk	4	76	0	0	321	13560	7	187
Segmental1	rish	rIS	5	312	0	0	296	6157	17	440
	group mean		7.00	333.50	0.00	0.00	377.75	17103.75	11.50	787.00
Segmental2	sove	s5v	13	97	6	586	1428	38166	18	127
Segmental2	samb	s{m	2	19	0	0	1428	38166	29	4711
Segmental2	jace	_1s	12	1251	0	0	104	2545	18	975
Segmental2	yate	j1t	16	865	1	40	71	10386	17	1128
	group mean		10.75	558.00	1.75	156.50	757.75	22315.75	20.50	1735.25
Unrelated1	virt	v3t	6	85	0	0	77	873	16	149
Unrelated1	musp	mVsp	1	0	0	0	296	20038	4	1095
Unrelated1	chen	JEn	18	5054	0	0	144	2420	22	7371
Unrelated1	galk	g{lk	0	0	5	362	389	10059	2	0
	group mean		6.25	1284.75	1.25	90.50	226.50	8347.50	11.00	2153.75
Unrelated2	fomb	fQm	1	28	3	27	447	24349	13	555
Unrelated2	cive	s2v	11	365	2	653	695	15539	19	1370
Unrelated2	kyth	kIT	1	19	0	0	88	3717	14	6379
Unrelated2	jasp	_{sp	0	0	6	31	104	2545	0	0
	group mean		3.25	103.00	2.75	177.75	333.50	11537.50	11.50	2076.00

group	spelling	pronunciation	bigram frequency (position nonspecific) - type	bigram frequency (position nonspecific) - token	trigram frequency (position nonspecific) - type	trigram frequency (position nonspecific) - token	bigram frequency (position specific) - type	bigram frequency (position specific) - token	trigram frequency (position specific) - type	trigram frequency (position specific) - token
Semantic1	chys	JIs	441	229046	1	10	38	22108	1	10
Semantic1	fulb	fVlb	131	121369	4	4881	33	15284	2	4790
Semantic1	vork	v9k	263	395848	13	20060	62	102915	5	16000
Semantic1	jing	_IN	514	540117	54	28937	84	64107	8	3779
	group mean		337.25	321595.00	18.00	13472.00	54.25	51103.50	4.00	6144.75
Semantic2	frop	frQp	525	312022	38	81723	39	156048	5	72771
Semantic2	wung	wVN	335	142673	27	11511	60	27935	6	1507
Semantic2	jalt	_{lt	262	159064	14	4846	68	43364	3	1114
Semantic2	byme	b2m	201	300332	4	158	23	115020	0	0
	group mean		330.75	228522.75	20.75	24559.50	47.50	85591.75	3.50	18848.00
Segmental1	fusp	fVsp	343	170043	6	475	48	46363	3	392
Segmental1	fing	fIN	561	595125	61	40339	99	86733	13	14343
Segmental1	lisk	lIsk	332	471694	17	3627	55	81091	3	2451
Segmental1	rish	rIS	644	582852	27	11647	69	33225	5	8217
	group mean		470.00	454928.50	27.75	14022.00	67.75	61853.00	6.00	6350.75
Segmental2	sove	s5v	337	284256	42	19665	57	169655	9	9398
Segmental2	samb	s{m	241	161642	5	11370	56	107906	3	11260
Segmental2	jace	_1s	288	182135	24	27442	50	77809	7	10914
Segmental2	yate	j1t	414	500960	31	22021	62	27203	9	9486
	group mean		320.00	282248.25	25.50	20124.50	56.25	95643.25	7.00	10264.50
Unrelated1	virt	v3t	207	158503	17	3193	36	25792	2	345
Unrelated1	musp	mVsp	352	198284	10	15553	52	75296	6	15501
Unrelated1	chen	JEn	857	2157153	119	91326	41	320628	4	73772
Unrelated1	galk	g{lk	270	168190	22	12031	66	50748	6	6396
	group mean		421.50	670532.50	42.00	30525.75	48.75	118116.00	4.50	24003.50
Unrelated2	fomb	fQm	163	330636	10	1493	45	76577	4	993
Unrelated2	cive	s2v	270	186675	36	31610	34	133771	8	20380
Unrelated2	kyth	kIT	290	1895416	5	498	13	131947	1	334
Unrelated2	jasp	_{sp	413	455963	16	1123	58	39428	4	145
	group mean		284.00	717172.50	16.75	8681.00	37.50	95430.75	4.25	5463.00

Table C2. Psycholinguistic characteristics of pseudoword stimuli used in the spoken experiment. Values are from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002).

group	spelling	pronunciation	letters	phonemes	orthographic neighbors	summed frequency of orthographic neighbors	body neighbors	summed frequency of body neighbors
Semantic1	crube	krub	5	4	1	20	2	19
Semantic1	dwipe	dw2p	5	4	1	1	8	45
Semantic1	haint	h1nt	5	4	5	85	5	81
Semantic1	smuth	smVT	5	4	3	228	2	123
	group mean		5.00	4.00	2.50	83.50	4.25	67.00
Semantic2	drage	dr1_	5	4	3	5	8	493
Semantic2	glybe	gl2b	5	4	2	10	1	0
Semantic2	skeph	skEf	5	4	2	1	1	1
Semantic2	shump	SVmp	5	4	4	16	16	105
	group mean		5.00	4.00	2.75	8.00	6.50	149.75
Segmental1	bleck	blEk	5	4	4	373	9	168
Segmental1	bline	bl2n	5	4	4	41	22	624
Segmental1	brudd	brVd	5	4	0	0	1	0
Segmental1	gluth	glVT	5	4	1	0	2	123
	group mean		5.00	4.00	2.25	103.50	8.50	228.75
Segmental2	swush	swVS	5	4	3	4	16	167
Segmental2	pluff	plVf	5	4	2	8	15	83
Segmental2	slaim	sl1m	5	4	2	72	3	110
Segmental2	sloot	slut	5	4	3	23	11	169
	group mean		5.00	4.00	2.50	26.75	11.25	132.25
Unrelated1	twish	twIS	5	4	3	14	5	312
Unrelated1	glafe	gl1f	5	4	3	11	3	77
Unrelated1	scoon	skun	5	4	5	26	10	399
Unrelated1	themp	TEmp	5	4	2	30	2	2
	group mean		5.00	4.00	3.25	20.25	5.00	197.50
Unrelated2	grive	gr2v	5	4	6	152	13	1018
Unrelated2	dwesh	dwES	5	4	0	0	4	126
Unrelated2	bloof	bluf	5	4	2	145	8	78
Unrelated2	thunt	TVnt	5	4	2	1	9	50
	group mean		5.00	4.00	2.50	74.50	8.50	318.00

group	spelling	pronunciation	body friends	summed frequency of body friends	body enemies	summed frequency of body enemies	onset neighbors	summed frequency of onset neighbors	phonological neighbors	summed frequency of phonological neighbors
Semantic1	crube	krub	2	19	0	0	135	873	9	39
Semantic1	dwipe	dw2p	8	45	0	0	7	20	1	0
Semantic1	haint	h1nt	5	81	0	0	301	41835	9	77
Semantic1	smuth	smVT	0	0	2	123	43	1012	4	61
	group mean		3.75	36.25	0.50	30.75	121.50	10935.00	5.75	44.25
Semantic2	drage	dr1_	7	492	1	1	90	1361	6	15
Semantic2	glybe	gl2b	1	0	0	0	48	399	4	4
Semantic2	skeph	skEf	1	1	0	0	46	374	8	22
Semantic2	shump	SVmp	16	105	0	0	175	7516	10	58
	group mean		6.25	149.50	0.25	0.25	89.75	2412.50	7.00	24.75
Segmental1	bleck	blEk	9	168	0	0	96	1122	8	265
Segmental1	bline	bl2n	22	624	0	0	96	1122	5	14
Segmental1	brudd	brVd	1	0	0	0	128	1661	12	147
Segmental1	gluth	glVT	0	0	2	123	48	399	3	2
	group mean		8.00	198.00	0.50	30.75	92.00	1076.00	7.00	107.00
Segmental2	swush	swVS	13	82	3	85	87	447	4	3
Segmental2	pluff	plVf	15	83	0	0	94	1857	10	299
Segmental2	slaim	sl1m	3	110	0	0	120	655	13	115
Segmental2	sloot	slut	9	57	2	112	120	655	13	47
	group mean		10.00	83.00	1.25	49.25	105.25	903.50	10.00	116.00
Unrelated1	twish	twIS	5	312	0	0	39	1496	7	14
Unrelated1	glafe	gl1f	2	77	1	0	48	399	2	1
Unrelated1	scoon	skun	10	399	0	0	135	1127	10	878
Unrelated1	themp	TEmp	2	2	0	0	119	95453	3	3
	group mean		4.75	197.50	0.25	0.00	85.25	24618.75	5.50	224.00
Unrelated2	grive	gr2v	11	365	2	653	129	2590	10	114
Unrelated2	dwesh	dwES	4	126	0	0	7	20	1	4
Unrelated2	bloof	bluf	8	78	0	0	96	1122	4	6
Unrelated2	thunt	TVnt	9	50	0	0	119	95453	5	12
	group mean		8.00	154.75	0.50	163.25	87.75	24796.25	5.00	34.00

group	spelling	pronunciation	bigram frequency (position nonspecific) - type	bigram frequency (position nonspecific) - token	trigram frequency (position nonspecific) - type	trigram frequency (position nonspecific) - token	bigram frequency (position specific) - type	bigram frequency (position specific) - token	trigram frequency (position specific) - type	trigram frequency (position specific) - token
Semantic1	crube	krub	499	230236	38	3370	89	20740	5	656
Semantic1	dwipe	dw2p	509	245097	25	2160	50	6863	5	57
Semantic1	haint	h1nt	885	1189608	121	38697	165	76253	24	11783
Semantic1	smuth	smVT	469	2112656	23	12136	78	58470	8	11417
	group mean		590.50	944399.25	51.75	14090.75	95.50	40581.50	10.50	5978.25
Semantic2	drage	dr1_	722	169749	54	16584	167	48610	14	4862
Semantic2	glybe	gl2b	198	184062	1	1	25	3997	0	0
Semantic2	skeph	skEf	444	182107	18	12168	43	5966	5	29
Semantic2	shump	SVmp	579	186911	70	5825	116	29575	17	1181
	group mean		485.75	180707.25	35.75	8644.50	87.75	22037.00	9.00	1518.00
Segmental1	bleck	blEk	757	232707	46	5647	121	46473	11	1851
Segmental1	bline	bl2n	872	744893	114	26736	158	64062	26	2808
Segmental1	brudd	brVd	319	66895	18	1549	73	23198	5	794
Segmental1	gluth	glVT	563	2078676	29	12429	119	50086	12	11455
	group mean		627.75	780792.75	51.75	11590.25	117.75	45954.75	13.50	4227.00
Segmental2	swush	swVS	476	296895	28	5744	85	34396	8	1124
Segmental2	pluff	plVf	357	68278	67	4063	91	23537	16	1857
Segmental2	slaim	sl1m	680	309622	65	11056	140	47836	22	4392
Segmental2	sloot	slut	734	389022	87	35863	138	37272	17	6795
	group mean		561.75	265954.25	61.75	14181.50	113.50	35760.25	15.75	3542.00
Unrelated1	twish	twIS	508	717602	40	13652	80	14468	13	2745
Unrelated1	glafe	gl1f	447	167762	23	6372	97	38522	6	2447
Unrelated1	scoon	skun	706	503434	76	12335	87	19839	17	1505
Unrelated1	themp	TEmp	769	3709943	77	1352392	90	334901	8	125472
	group mean		607.50	1274685.25	54.00	346187.75	88.50	101932.50	11.00	33042.25
Unrelated2	grive	gr2v	669	307439	75	35822	159	67537	15	3677
Unrelated2	dwesh	dwES	964	502465	14	5689	59	37364	4	2275
Unrelated2	bloof	bluf	626	784711	75	35646	128	45526	19	10628
Unrelated2	thunt	TVnt	691	2041255	63	9901	130	235086	21	2956
	group mean		737.50	908967.50	56.75	21764.50	119.00	96378.25	14.75	4884.00

Appendix D: Tables of Results from Study 1

Table D1. Results of the analysis of the semantic model of the written training data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	3.35	0.49	6.88	<.001	7.02	0.04	198.07	<.001
block type (semantic vs. unrelated)	-0.69	0.33	-2.08	.038	0.02	0.02	0.93	.350
training attempt within session	1.94	0.32	6.11	<.001	-0.09	0.01	-6.53	<.001
session	2.26	0.30	7.43	<.001	-0.23	0.01	-16.24	<.001
days since last session	-0.03	0.17	-0.17	.862	0.02	0.01	1.25	.210
training trials since last trained	-0.13	0.13	-1.01	.311	-0.02	0.01	-2.12	.034
block type (semantic vs. unrelated) * training attempt within session	-0.57	0.26	-2.21	.027	0.01	0.01	1.22	.221
block type (semantic vs. unrelated) * session	-0.44	0.23	-1.96	.050	-0.01	0.01	-1.41	.159
training attempt within session * session	-0.21	0.25	-0.85	.398	0.02	0.01	2.63	.008
block type (semantic vs. unrelated) * training attempt within session * session	-0.38	0.22	-1.74	.081	-0.02	0.01	-1.97	.049

Random effects	Variance	Variance
subject intercept	2.1779	0.0120
block type (semantic vs. unrelated)|subject slope	0.2059	0.0005
training attempt within session|subject slope	0.2392	0.0013
session|subject slope	0.6689	0.0025
days since last session|subject slope	0.1202	0.0011
training trials since last trained|subject slope	0.0692	0.0004
block type (semantic vs. unrelated) * training attempt within session|subject slope	0.2369	0.0003
block type (semantic vs. unrelated) * session|subject	0.0662	0.0009

slope		
training attempt within session * session|subject slope	0.1289	0.0002
block type (semantic vs. unrelated) * training attempt within session * session|subject slope	0.1154	0.0002
item intercept	0.6547	0.0082
training attempt within session|item slope	0.0684	0.0004
session|item slope	0.1061	0.0003
days since last session |item slope	0.0182	0.0009
training trials since last trained|item slope	0.0279	0.0001
training attempt within session * session|item slope	0.0808	0.0001
Residual		0.0869

Table D2. Results of the analysis of the segmental model of the written training data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	3.66	0.53	6.92	<.001	6.99	0.04	191.36	<.001
block type (segmental vs. unrelated)	-0.60	0.33	-1.81	.070	-0.01	0.03	-0.36	.716
training attempt within session	2.13	0.42	5.10	<.001	-0.07	0.01	-5.91	<.001
session	2.48	0.31	8.00	<.001	-0.21	0.01	-15.94	<.001
days since last session	-0.13	0.14	-0.88	.378	0.02	0.01	1.03	.304
training trials since last trained	0.07	0.12	0.61	.545	-0.02	0.02	-1.24	.216
block type (segmental vs. unrelated) * training attempt within session	-0.42	0.28	-1.54	.125	0.02	0.01	3.20	.001
block type (segmental vs. unrelated) * session	-0.30	0.23	-1.31	.190	0.01	0.01	0.91	.364
training attempt within session * session	-0.03	0.33	-0.10	.919	0.01	0.01	1.30	.194
block type (segmental vs. unrelated) * training attempt within session * session	-0.23	0.24	-0.99	.321	-0.02	0.01	-2.57	.010

Random effects	Variance	Variance
subject intercept	2.6845	0.0102
block type (segmental vs. unrelated)|subject slope	0.0796	0.0004
training attempt within session|subject slope	0.9401	0.0010
session|subject slope	0.5335	0.0011
days since last session|subject slope	0.0041	0.0011
training trials since last trained|subject slope	0.0143	0.0010
block type (segmental vs. unrelated) * training attempt within session|subject slope	0.0338	0.0001
block type (segmental vs. unrelated) * session|subject slope	0.0343	0.0006
training attempt within session * session|subject slope	0.5874	0.0007

block type (segmental vs. unrelated) * training attempt within session * session|subject slope	0.0089	0.0001
item intercept	0.4328	0.0111
training attempt within session|item slope	0.1303	0.0002
session|item slope	0.0486	0.0009
days since last session |item slope	0.0057	0.0006
training trials since last trained|item slope	0.0142	0.0008
training attempt within session * session|item slope	0.1853	0.0004
Residual		0.0969

Table D3. Results of the analysis of the semantic model of the spoken training data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	2.30	0.45	5.05	<.001	6.99	0.07	107.29	<.001
block type (semantic vs. unrelated)	-0.34	0.26	-1.29	.198	0.03	0.03	0.95	.343
training attempt within session	1.64	0.23	7.16	<.001	-0.07	0.02	-3.36	.001
session	1.93	0.18	10.78	<.001	-0.18	0.02	-7.43	<.001
days since last session	-0.22	0.11	-1.93	.054	0.02	0.02	1.11	.265
training trials since last trained	-0.11	0.12	-0.87	.386	0.02	0.02	1.18	.237
block type (semantic vs. unrelated) * training attempt within session	-0.12	0.14	-0.85	.393	0.03	0.01	2.34	.019
block type (semantic vs. unrelated) * session	-0.05	0.16	-0.31	.756	-0.03	0.01	-2.31	.021
training attempt within session * session	-0.37	0.16	-2.39	.017	0.01	0.01	1.00	.319
block type (semantic vs. unrelated) * training attempt within session * session	-0.03	0.15	-0.21	.835	-0.01	0.01	-1.07	.283

Random effects	Variance	Variance
subject intercept	2.2875	0.0564
block type (semantic vs. unrelated)|subject slope	0.0962	0.0013
training attempt within session|subject slope	0.2467	0.0024
session|subject slope	0.1822	0.0068
days since last session|subject slope	0.0474	0.0011
training trials since last trained|subject slope	0.0833	0.0025
block type (semantic vs. unrelated) * training attempt within session|subject slope	0.0215	0.0003
block type (semantic vs. unrelated) * session|subject slope	0.1112	0.0004
training attempt within session * session|subject slope	0.0176	0.0015

block type (semantic vs. unrelated) * training attempt within session * session|subject slope	0.0794	0.0009
item intercept	0.7950	0.0137
training attempt within session|item slope	0.1037	0.0019
session|item slope	0.0601	0.0014
days since last session |item slope	0.0225	0.0006
training trials since last trained|item slope	0.0255	0.0007
training attempt within session * session|item slope	0.0685	0.0009
Residual		0.1236

Table D4. Results of the analysis of the segmental model of the spoken training data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	2.48	0.44	5.63	<.001	6.98	0.06	113.39	<.001
block type (segmental vs. unrelated)	-0.08	0.29	-0.29	.773	0.00	0.03	0.13	.896
training attempt within session	1.72	0.22	7.79	<.001	-0.09	0.02	-5.10	<.001
session	1.99	0.18	10.99	<.001	-0.16	0.02	-6.51	<.001
days since last session	-0.15	0.12	-1.27	.205	0.04	0.02	2.43	.015
training trials since last trained	-0.10	0.10	-0.97	.331	-0.02	0.02	-1.08	.279
block type (segmental vs. unrelated) * training attempt within session	-0.15	0.16	-0.92	.355	0.02	0.01	2.00	.046
block type (segmental vs. unrelated) * session	0.01	0.15	0.08	.935	-0.01	0.01	-1.17	.242
training attempt within session * session	-0.31	0.15	-1.99	.046	0.01	0.01	0.62	.534
block type (segmental vs. unrelated) * training attempt within session * session	-0.07	0.14	-0.49	.623	-0.02	0.01	-1.61	.108

Random effects	Variance	Variance
subject intercept	1.7126	0.0508
block type (segmental vs. unrelated)|subject slope	0.1208	0.0021
training attempt within session|subject slope	0.1621	0.0014
session|subject slope	0.1990	0.0086
days since last session|subject slope	0.0436	0.0014
training trials since last trained|subject slope	0.0107	0.0010
block type (segmental vs. unrelated) * training attempt within session|subject slope	0.0269	0.0001
block type (segmental vs. unrelated) * session|subject slope	0.0561	0.0002
training attempt within session * session|subject slope	0.0232	0.0008

block type (segmental vs. unrelated) * training attempt within session * session|subject slope	0.0303	0.0003
item intercept	1.1202	0.0117
training attempt within session|item slope	0.1066	0.0017
session|item slope	0.0473	0.0002
days since last session |item slope	0.0200	0.0000
training trials since last trained|item slope	0.0154	0.0006
training attempt within session * session|item slope	0.0183	0.0012
Residual		0.1261

Table D5. Results of the analysis of the model of the written semantic probe over sessions data comparing shared and distinctive features trained in semantic blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.39	0.14	9.59	<.001	6.82	0.03	259.00	<.001
feature type (distinctive vs. shared)	0.06	0.07	0.79	.429	-0.02	0.01	-2.84	.005
session	0.15	0.08	1.96	.050	-0.04	0.01	-5.11	<.001
days since last session	0.05	0.07	0.71	.480	0.00	0.01	0.43	.669
feature type (distinctive vs. shared) * session	0.08	0.07	1.13	.261	-0.01	0.01	-1.72	.085

Random effects	Variance	Variance
subject intercept	0.1334	0.0109
feature type (distinctive vs. shared)|subject slope	0.0341	0.0004
session|subject slope	0.0240	0.0006
days since last session|subject slope	0.0050	0.0007
feature type (distinctive vs. shared) * session|subject slope	0.0254	0.0000
item intercept	0.0793	0.0002
Residual		0.0432

Table D6. Results of the analysis of the model of the written semantic probe over sessions data comparing shared features trained in semantic blocks to distinctive features trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.77	0.13	13.24	<.001	6.79	0.03	250.44	<.001
feature type (distinctive vs. shared)	0.40	0.09	4.23	<.001	-0.06	0.01	-4.97	<.001
session	0.26	0.09	2.86	.004	-0.04	0.01	-6.20	<.001
days since last session	0.12	0.07	1.65	.099	0.00	0.01	-0.27	.790
feature type (distinctive vs. shared) * session	0.21	0.07	3.27	.001	-0.01	0.00	-2.57	.010

Random effects	Variance	Variance
item intercept	0.1105	0.0020
subject intercept	0.1727	0.0107
feature type (distinctive vs. shared)|subject slope	0.0235	0.0005
session|subject slope	0.0753	0.0005
days since last session|subject slope	0.0174	0.0003
feature type (distinctive vs. shared) * session|subject slope	0.0352	0.0001
Residual		0.0370

Table D7. Results of the analysis of the model of the spoken semantic probe over sessions data comparing shared and distinctive features trained in semantic blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.59	0.17	9.56	<.001	6.86	0.03	225.77	<.001
feature type (distinctive vs. shared)	0.14	0.09	1.54	.124	-0.04	0.01	-6.01	<.001
session	0.27	0.10	2.78	.005	-0.02	0.01	-1.90	.058
days since last session	0.09	0.12	0.71	.480	0.01	0.01	0.90	.368
feature type (distinctive vs. shared) * session	0.18	0.06	2.95	.003	-0.01	0.01	-2.24	.025

Random effects	Variance	Variance
subject intercept	0.2755	0.0144
feature type (distinctive vs. shared)|subject slope	0.0839	0.0003
session|subject slope	0.0560	0.0019
days since last session|subject slope	0.1159	0.0010
feature type (distinctive vs. shared) * session|subject slope	0.0003	0.0001
item intercept	0.0536	0.0003
Residual		0.0489

Table D8. Results of the analysis of the model of the spoken semantic probe over sessions data comparing shared features trained in semantic blocks to distinctive features trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.99	0.17	11.61	<.001	6.82	0.03	227.65	<.001
feature type (distinctive vs. shared)	0.53	0.13	4.25	<.001	-0.08	0.01	-6.32	<.001
session	0.32	0.08	3.87	<.001	-0.03	0.01	-2.56	.010
days since last session	0.06	0.10	0.56	.573	0.00	0.01	0.06	.950
feature type (distinctive vs. shared) * session	0.24	0.06	4.01	<.001	-0.02	0.01	-3.49	<.001

Random effects	Variance	Variance
item intercept	0.1994	0.0020
subject intercept	0.2822	0.0133
feature type (distinctive vs. shared)|subject slope	0.0585	0.0008
session|subject slope	0.0393	0.0013
days since last session|subject slope	0.0615	0.0013
feature type (distinctive vs. shared) * session|subject slope	0.0187	0.0002
Residual		0.0430

Table D9. Results of the analysis of the model of the written segment probe over sessions data comparing shared and distinctive segments trained in segmental blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.32	0.25	5.30	<.001	6.76	0.03	207.53	<.001
feature type (distinctive vs. shared)	-0.13	0.07	-1.78	.075	0.01	0.01	2.27	.023
session	0.74	0.11	6.88	<.001	-0.04	0.02	-2.02	.044
days since last session	0.06	0.08	0.75	.451	0.00	0.01	-0.21	.834
feature type (distinctive vs. shared) * session	-0.12	0.07	-1.72	.085	-0.01	0.01	-1.40	.160

Random effects	Variance	Variance
subject intercept	0.5841	0.0137
feature type (distinctive vs. shared)|subject slope	0.0168	0.0000
session|subject slope	0.0813	0.0059
days since last session|subject slope	0.0005	0.0011
feature type (distinctive vs. shared) * session|subject slope	0.0123	0.0001
item intercept	0.1845	0.0017
Residual		0.0562

Table D10. Results of the analysis of the model of the written segment probe over sessions data comparing shared segments trained in segmental blocks to distinctive segments trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.33	0.25	5.22	<.001	6.76	0.03	211.77	<.001
feature type (distinctive vs. shared)	-0.14	0.15	-0.92	.357	0.02	0.01	1.25	.212
session	0.77	0.10	7.68	<.001	-0.04	0.02	-1.76	.078
days since last session	0.06	0.06	0.94	.346	0.00	0.01	-0.07	.941
feature type (distinctive vs. shared) * session	-0.10	0.07	-1.54	.123	0.00	0.01	-0.84	.401

Random effects	Variance	Variance
item intercept	0.3842	0.0026
subject intercept	0.7431	0.0149
feature type (distinctive vs. shared)|subject slope	0.0395	0.0002
session|subject slope	0.0921	0.0070
days since last session|subject slope	0.0083	0.0011
feature type (distinctive vs. shared) * session|subject slope	0.0256	0.0002
Residual		0.0584

Table D11. Results of the analysis of the model of the spoken segment probe over sessions data comparing shared and distinctive segments trained in segmental blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.11	0.29	3.77	<.001	7.17	0.04	169.13	<.001
feature type (distinctive vs. shared)	-0.02	0.07	-0.33	.743	0.00	0.01	-0.46	.644
session	0.74	0.11	6.80	<.001	-0.05	0.02	-2.79	.005
days since last session	-0.01	0.12	-0.05	.958	0.03	0.01	2.15	.032
feature type (distinctive vs. shared) * session	0.00	0.07	0.04	.969	0.01	0.01	1.92	.055

Random effects	Variance	Variance
subject intercept	0.9367	0.0246
feature type (distinctive vs. shared)|subject slope	0.0293	0.0005
session|subject slope	0.0876	0.0041
days since last session|subject slope	0.1090	0.0008
feature type (distinctive vs. shared) * session|subject slope	0.0181	0.0001
item intercept	0.2096	0.0024
Residual		0.0624

Table D12. Results of the analysis of the model of the spoken segment probe over sessions data comparing shared segments trained in segmental blocks to distinctive segments trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.05	0.24	4.43	<.001	7.16	0.04	202.26	<.001
feature type (distinctive vs. shared)	-0.12	0.12	-1.00	.316	-0.01	0.01	-0.76	.448
session	0.60	0.07	8.56	<.001	-0.06	0.02	-3.44	.001
days since last session	0.10	0.08	1.30	.193	0.03	0.01	3.31	.001
feature type (distinctive vs. shared) * session	-0.09	0.05	-1.79	.073	0.01	0.01	1.83	.067

Random effects	Variance	Variance
item intercept	0.1782	0.0022
subject intercept	0.7642	0.0191
feature type (distinctive vs. shared)|subject slope	0.0609	0.0000
session|subject slope	0.0272	0.0043
days since last session|subject slope	0.0424	0.0011
feature type (distinctive vs. shared) * session|subject slope	0.0023	0.0000
Residual		0.0632

Table D13. Results of the analysis of the model of the written semantic probe follow-up data comparing shared and distinctive features trained in semantic blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.26	0.22	5.85	<.001	6.82	0.03	229.89	<.001
feature type (distinctive vs. shared)	0.06	0.11	0.52	.601	-0.02	0.01	-1.87	.061
Random effects	Variance				Variance			
subject intercept	0.3572				0.0130			
feature type (distinctive vs. shared)|subject slope	0.0049				0.0000			
item intercept	0.1042				0.0000			
Residual					0.0445			

Table D14. Results of the analysis of the model of the written semantic probe follow-up data comparing shared features trained in semantic blocks to distinctive features trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.69	0.20	8.48	<.001	6.80	0.03	230.84	<.001
feature type (distinctive vs. shared)	0.46	0.11	4.01	<.001	-0.04	0.01	-3.90	<.001
Random effects	Variance				Variance			
item intercept	0.0855				0.0012			
subject intercept	0.4459				0.0127			
feature type (distinctive vs. shared)|subject slope	0.0057				0.0001			
Residual					0.0385			

Table D15. Results of the analysis of the model of the spoken semantic probe follow-up data comparing shared and distinctive features trained in semantic blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.24	0.12	10.54	<.001	6.89	0.03	197.41	<.001
feature type (distinctive vs. shared)	0.20	0.11	1.86	.062	-0.01	0.01	-1.14	.252
days since last session	-0.03	0.13	-0.21	.834	-0.06	0.02	-2.67	.008

Random effects	Variance (Accuracy)	Variance (RT)
subject intercept	0.0294	0.0146
feature type (distinctive vs. shared)|subject slope	0.0134	0.0002
days since last session|subject slope	0.0251	0.0031
item intercept	0.0000	0.0012
Residual		0.0509

Table D16. Results of the analysis of the model of the spoken semantic probe follow-up data comparing shared features trained in semantic blocks to distinctive features trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.74	0.19	9.39	<.001	6.84	0.03	223.37	<.001
feature type (distinctive vs. shared)	0.63	0.13	4.72	<.001	-0.06	0.02	-3.60	<.001
days since last session	-0.18	0.20	-0.92	.358	-0.01	0.03	-0.53	.595
Random effects	Variance				Variance			
item intercept	0.1484				0.0024			
subject intercept	0.2526				0.0087			
feature type (distinctive vs. shared)|subject slope	0.0382				0.0011			
days since last session|subject slope	0.0378				0.0082			
Residual					0.0504			

Table D17. Results of the analysis of the model of the written segment probe follow-up data comparing shared and distinctive segments trained in segmental blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.69	0.35	4.80	<.001	6.73	0.03	240.44	<.001
feature type (distinctive vs. shared)	-0.40	0.14	-2.81	.005	0.01	0.01	1.27	.202

Random effects	Variance	Variance
subject intercept	0.6190	0.0090
feature type (distinctive vs. shared)|subject slope	0.0528	0.0001
item intercept	0.5351	0.0011
Residual		0.0482

Table D18. Results of the analysis of the model of the written segment probe follow-up data comparing shared segments trained in segmental blocks to distinctive segments trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.71	0.30	5.69	<.001	6.73	0.03	241.52	<.001
feature type (distinctive vs. shared)	-0.37	0.22	-1.69	.092	0.02	0.02	1.03	.301

Random effects	Variance	Variance
item intercept	0.5572	0.0038
subject intercept	0.8369	0.0091
feature type (distinctive vs. shared)|subject slope	0.1201	0.0001
Residual		0.0459

Table D19. Results of the analysis of the model of the spoken segment probe follow-up data comparing shared and distinctive segments trained in segmental blocks, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.34	0.22	6.04	<.001	7.13	0.05	144.11	<.001
feature type (distinctive vs. shared)	0.11	0.13	0.84	.400	0.02	0.01	1.05	.293
days since last session	0.67	0.21	3.27	.001	0.01	0.04	0.14	.886

Random effects	Variance	Variance
subject intercept	0.3429	0.0354
feature type (distinctive vs. shared)|subject slope	0.0532	0.0010
days since last session|subject slope	0.0250	0.0004
item intercept	0.0716	0.0016
Residual		0.0580

Table D20. Results of the analysis of the model of the spoken segment probe follow-up data comparing shared segments trained in segmental blocks to distinctive segments trained in other contexts, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.24	0.23	5.41	<.001	7.11	0.04	177.99	<.001
feature type (distinctive vs. shared)	-0.05	0.15	-0.36	.723	0.01	0.02	0.51	.607
days since last session	0.56	0.22	2.51	.012	-0.02	0.04	-0.48	.631

Random effects	Variance	Variance
item intercept	0.2618	0.0024
subject intercept	0.3533	0.0136
feature type (distinctive vs. shared)|subject slope	0.0527	0.0009
days since last session|subject slope	0.0955	0.0160
Residual		0.0600

Table D21. Results of the analysis of the semantic model of the written recall over sessions data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.82	0.54	3.38	.001	7.26	0.05	138.62	<.001
block type (semantic vs. unrelated)	-0.04	0.31	-0.15	.885	0.01	0.04	0.25	.801
relative time within recall episode (earlier vs. later)	-0.39	0.14	-2.81	.005	0.06	0.02	3.61	<.001
recall episode	1.65	0.16	10.49	<.001	-0.20	0.02	-8.67	<.001
block type (semantic vs. unrelated) * relative time within recall episode	0.04	0.12	0.34	.737	0.00	0.01	0.27	.787
block type (semantic vs. unrelated) * recall episode	0.05	0.14	0.40	.693	0.00	0.01	-0.23	.818
relative time within recall episode * recall episode	-0.04	0.13	-0.31	.755	0.02	0.01	1.38	.167
block type (semantic vs. unrelated) * relative time within recall episode * recall episode	-0.02	0.12	-0.20	.840	-0.01	0.02	-0.32	.751

Random effects	Variance	Variance
subject intercept	3.3069	0.0218
block type (semantic vs. unrelated)|subject slope	0.0932	0.0005
relative time within recall episode (earlier vs. later) |subject slope	0.0690	0.0013
recall episode|subject slope	0.1294	0.0060
block type (semantic vs. unrelated) * relative time within recall episode|subject slope	0.0091	0.0002
block type (semantic vs. unrelated) * recall episode|subject slope	0.0548	0.0010
training attempt within session * recall episode|subject slope	0.0624	0.0011
block type (semantic vs. unrelated) *relative time within	0.0058	0.0027

recall episode* recall episode|subject slope		
item intercept	1.1536	0.0207
Residual		0.1691

Table D22. Results of the analysis of the segmental model of the written recall over sessions data, including both accuracy and response time

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.81	0.51	3.57	<.001	7.25	0.05	133.92	<.001
block type (segmental vs. unrelated)	-0.10	0.34	-0.30	.768	-0.01	0.04	-0.24	.810
relative time within recall episode (earlier vs. later)	-0.40	0.12	-3.22	.001	0.06	0.02	3.69	<.001
recall episode	1.40	0.14	9.87	<.001	-0.20	0.02	-9.40	<.001
block type (segmental vs. unrelated) * relative time within recall episode	0.03	0.11	0.29	.773	0.01	0.01	0.91	.364
block type (segmental vs. unrelated) * recall episode	-0.25	0.13	-1.92	.055	-0.01	0.01	-0.71	.477
relative time within recall episode * recall episode	-0.04	0.12	-0.37	.714	0.01	0.01	0.95	.341
block type (segmental vs. unrelated) * relative time within recall episode * recall episode	-0.03	0.12	-0.29	.769	-0.01	0.02	-0.74	.457
Random effects	Variance				Variance			
subject intercept	2.6950				0.0192			
block type (segmental vs. unrelated)|subject slope	0.3308				0.0002			
relative time within recall episode (earlier vs. later) |subject slope	0.0315				0.0020			
recall episode|subject slope	0.0778				0.0051			
block type (segmental vs. unrelated) * relative time within recall episode|subject slope	0.0032				0.0003			
block type (segmental vs. unrelated) * recall episode|subject slope	0.0439				0.0008			
training attempt within session * recall episode|subject slope	0.0363				0.0006			
block type (segmental vs. unrelated) *relative time	0.0276				0.0022			

within recall episode* recall episode|subject slope		
item intercept	1.2472	0.0260
Residual		0.1831

Table D23. Results of the analysis of the semantic model of the spoken recall over sessions data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	0.66	0.44	1.49	.137	7.25	0.08	94.90	<.001
block type (semantic vs. unrelated)	-0.02	0.27	-0.06	.951	0.01	0.05	0.14	.888
relative time within recall episode (earlier vs. later)	-0.34	0.09	-3.58	<.001	0.04	0.02	2.16	.031
recall episode	1.36	0.13	10.80	<.001	-0.16	0.02	-6.79	<.001
block type (semantic vs. unrelated) * relative time within recall episode	-0.06	0.08	-0.75	.453	0.00	0.02	0.30	.767
block type (semantic vs. unrelated) * recall episode	0.10	0.09	1.16	.246	0.00	0.02	-0.27	.786
relative time within recall episode * recall episode	0.03	0.09	0.38	.702	-0.01	0.02	-0.32	.751
block type (semantic vs. unrelated) * relative time within recall episode * recall episode	0.01	0.09	0.16	.870	0.01	0.02	0.62	.536

Random effects	Variance	Variance
subject intercept	2.2823	0.0566
block type (semantic vs. unrelated)|subject slope	0.1277	0.0006
relative time within recall episode (earlier vs. later) |subject slope	0.0458	0.0013
recall episode|subject slope	0.1401	0.0049
block type (semantic vs. unrelated) * relative time within recall episode|subject slope	0.0166	0.0005
block type (semantic vs. unrelated) * recall episode|subject slope	0.0131	0.0010
training attempt within session * recall episode|subject slope	0.0142	0.0008
block type (semantic vs. unrelated) *relative time within	0.0211	0.0009

recall episode* recall episode|subject slope		
item intercept	0.9016	0.0349
Residual		0.1948

Table D24. Results of the analysis of the segmental model of the spoken recall over sessions data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	0.74	0.41	1.80	.073	7.26	0.08	91.72	<.001
block type (segmental vs. unrelated)	0.08	0.31	0.26	.797	0.01	0.06	0.19	.850
relative time within recall episode (earlier vs. later)	-0.32	0.09	-3.46	.001	0.02	0.02	1.44	.149
recall episode	1.30	0.10	12.72	<.001	-0.17	0.02	-8.72	<.001
block type (segmental vs. unrelated) * relative time within recall episode	-0.04	0.08	-0.47	.637	-0.01	0.02	-0.57	.567
block type (segmental vs. unrelated) * recall episode	0.04	0.09	0.40	.691	-0.01	0.02	-0.37	.711
relative time within recall episode * recall episode	0.05	0.08	0.64	.521	-0.01	0.02	-0.36	.721
block type (segmental vs. unrelated) * relative time within recall episode * recall episode	0.04	0.08	0.50	.620	0.01	0.01	0.71	.476

Random effects	Variance	Variance
subject intercept	1.4326	0.0558
block type (segmental vs. unrelated)|subject slope	0.1912	0.0032
relative time within recall episode (earlier vs. later) |subject slope	0.0493	0.0009
recall episode|subject slope	0.0604	0.0022
block type (segmental vs. unrelated) * relative time within recall episode|subject slope	0.0131	0.0006
block type (segmental vs. unrelated) * recall episode|subject slope	0.0363	0.0029
training attempt within session * recall episode|subject slope	0.0145	0.0005
block type (segmental vs. unrelated) *relative time	0.0212	0.0001

within recall episode* recall episode|subject slope		
item intercept	1.2714	0.0432
Residual		0.1911

Table D25. Results of the analysis of the semantic model of the written recall follow-up data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.52	0.41	3.71	<.001	7.50	0.06	115.98	<.001
block type (semantic vs. unrelated)	0.04	0.31	0.13	.897	0.06	0.05	1.43	.151
Random effects	Variance				Variance			
subject intercept	1.2471				0.0359			
block type (semantic vs. unrelated)|subject slope	0.1150				0.0003			
item intercept	0.8845				0.0204			
Residual					0.1446			

Table D26. Results of the analysis of the segmental model of the written recall follow-up data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.99	0.50	4.01	<.001	7.39	0.06	118.69	<.001
block type (segmental vs. unrelated)	0.53	0.41	1.29	.196	-0.02	0.05	-0.32	.749
Random effects	Variance				Variance			
subject intercept	1.3977				0.0323			
block type (segmental vs. unrelated)|subject slope	0.4005				0.0114			
item intercept	1.1838				0.0020			
Residual					0.3719			

Table D27. Results of the analysis of the semantic model of the spoken recall follow-up data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.16	0.36	3.23	.001	7.47	0.08	96.16	<.001
block type (semantic vs. unrelated)	-0.04	0.27	-0.17	.869	-0.10	0.04	-2.21	.027
days since last session	0.51	0.42	1.20	.230	-0.01	0.07	-0.17	.867
Random effects	Variance				Variance			
subject intercept	0.7825				0.0502			
block type (semantic vs. unrelated)|subject slope	0.1557				0.0008			
days since last session |subject slope	0.2998				0.0045			
item intercept	0.6035				0.0236			
days since last session |item slope	0.0014				0.0292			
Residual					0.2260			

Table D28. Results of the analysis of the segmental model of the spoken recall follow-up data, including both accuracy and response time.

	Accuracy				RT			
Fixed effects	Coefficient	SE	*z*	*p*	Coefficient	SE	*t*	*p*
Intercept	1.26	0.28	4.52	<.001	7.50	0.07	106.25	<.001
block type (segmental vs. unrelated)	0.20	0.24	0.85	.393	-0.03	0.04	-0.76	.445
days since last session	0.44	0.29	1.53	.125	-0.02	0.08	-0.24	.812

Random effects	Variance	Variance
subject intercept	0.2992	0.0447
block type (segmental vs. unrelated)|subject slope	0.0001	0.0014
days since last session |subject slope	0.0463	0.0116
item intercept	0.4697	0.0074
days since last session |item slope	0.0224	0.0140
Residual		0.2339

Appendix E: Lists of Treatment Words for Each Participant in Study 2

Table E1. Treatment words for participant DDR.

	Treatment Context		
Set	Semantic	Orthographic	Unrelated
Set 1	soap	truck	ballet
	razor	clock	doctor
	toothpaste	lock	fish
	bath	neck	house
Set 2	brother	telephone	kitchen
	wife	cane	taxi
	uncle	line	dog
	friend	pine	gas
Set 3	clay	door	boat
	wire	hook	map
	plaster	book	email
	stone	spoon	table

Table E2. Treatment words for participant DWS.

	Treatment Context		
Set	Semantic	Orthographic	Unrelated
Set 1	tomato	blanket	theater
	orange	net	direction
	banana	rocket	chicken
	cherry	jacket	fork
Set 2	coffee	doctor	email
	soda	razor	fish
	water	anchor	hospital
	tea	mirror	music
Set 3	bed	house	golf
	table	mouse	map
	desk	nurse	thumb
	couch	rose	knife

Table E3. Treatment words for participant ESG.

	Treatment Context		
Set	Semantic	Orthographic	Unrelated
Set 1	mouse	table	bathroom
	computer	apple	driver
	email	bottle	cane
	phone	whistle	knife
Set 2	church	doctor	plate
	hospital	razor	tongue
	bank	mirror	snow
	garage	tractor	desk
Set 3	truck	watch	brain
	bicycle	speech	television
	plane	lunch	shower
	ship	bench	pencil

Table E4. Treatment words for participant GHN.

	Treatment Context		
Set	Semantic	Orthographic	Unrelated
Set 1	banana	table	landscape
	tomato	bottle	coat
	yeast	circle	bluejay
	sugar	ankle	phone
Set 2	hammer	horse	cane
	plane	birdhouse	basement
	drill	cheese	shirt
	clamp	vase	dishes
Set 3	deodorant	paint	kitchen
	tissue	brain	squirrel
	lotion	raisin	floor
	bandage	stairs	blanket

Table E5. Treatment words for participant KSR2.

	Treatment Context		
Set	Semantic	Orthographic	Unrelated
Set 1	sushi	doctor	father
	burger	razor	pants
	pasta	floor	table
	rice	anchor	golf
Set 2	phone	college	daughter
	television	judge	hospital
	email	page	music
	computer	bandage	couch
Set 3	leg	snail	crab
	hand	tail	ship
	brain	nail	wife
	mouth	pencil	movie

Table E6. Treatment words for participant REN.

Set	Treatment Context		
	Semantic	Orthographic	Unrelated
Set 1	track	razor	drama
	baseball	horror	mom
	golf	doctor	polish
	skating	anchor	apartment
Set 2	theater	racket	bicycle
	restaurant	vet	magazine
	gym	wallet	pants
	store	toilet	shampoo
Set 3	stomach	cash	television
	brain	flash	niece
	tooth	bush	war
	knee	trash	scarf

Table E7. Treatment words for participant SMY.

	Treatment Context		
Set	Semantic	Orthographic	Unrelated
Set 1	stadium	alphabet	husband
	exhibit	photograph	anchor
	cinema	telephone	mosaic
	orchestra	nephew	passport
Set 2	finger	speech	cash
	tongue	architect	violin
	elbow	watch	gallery
	mouth	porch	soap
Set 3	penguin	forest	ticket
	squirrel	restaurant	sculpture
	turkey	question	pants
	elephant	whistle	accordion

Appendix F: Tables of Results from Study 2

Table F1. Results of the analysis of the semantic model of the training data.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	2.85	0.42	6.75	<.001
block type (semantic vs. unrelated)	-0.06	0.12	-0.53	.596
training attempt within session	1.17	0.20	5.86	<.001
session	1.10	0.15	7.32	<.001
frequency (Kucera-Francis)	-0.07	0.09	-0.74	.458
length (letters)	-0.35	0.09	-3.93	<.001
days since last training session	-0.08	0.06	-1.31	.191
trials since last trained	-0.12	0.06	-2.11	.035
block type (semantic vs. unrelated)*training attempt within session	0.00	0.06	0.05	.963
block type (semantic vs. unrelated)*session	0.02	0.08	0.26	.797
training attempt within session*session	0.20	0.10	2.13	.033
block type (semantic vs. unrelated)*training attempt within session*session	0.01	0.06	0.11	.912

Random effects	Variance
item intercept	1.5855
training attempt within session|item slope	0.4568
session|item slope	0.3396
days since last training session|item slope	0.0088
trials since last trained|item slope	0.0378
training attempt within session*session|item slope	0.0965
subject intercept	1.0885
block type (semantic vs. unrelated)|subject slope	0.0554
training attempt within session|subject slope	0.1876
session|subject slope	0.0967
frequency (Kucera-Francis)|subject slope	0.0127
length (letters)|subject slope	0.0158
days since last training session|subject slope	0.0099
trials since last trained|subject slope	0.0071
block type (semantic vs. unrelated)*training attempt within session|subject slope	0.0059
block type (semantic vs. unrelated)*session|subject slope	0.0199
training attempt within session*session|subject slope	0.0180
block type (semantic vs. unrelated)*training attempt within session*session|subject slope	0.0095

Table F2. Results of the analysis of the segmental model of the training data.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	3.05	0.37	8.23	<.001
block type (segmental vs. unrelated)	0.56	0.16	3.54	<.001
training attempt within session	1.14	0.18	6.36	<.001
session	0.90	0.12	7.40	<.001
frequency (Kucera-Francis)	0.23	0.11	2.02	.043
length (letters)	-0.11	0.11	-1.01	.312
days since last training session	-0.11	0.07	-1.51	.132
trials since last trained	-0.19	0.07	-2.59	.010
block type (segmental vs. unrelated)*training attempt within session	0.23	0.08	2.69	.007
block type (segmental vs. unrelated)*session	0.05	0.09	0.57	.571
training attempt within session*session	-0.03	0.08	-0.41	.679
block type (segmental vs. unrelated)*training attempt within session*session	-0.08	0.05	-1.60	.110

Random effects	Variance
item intercept	1.7917
training attempt within session|item slope	0.3210
session|item slope	0.2589
days since last training session|item slope	0.0094
trials since last trained|item slope	0.0382
training attempt within session*session|item slope	0.0820
subject intercept	0.7966
block type (segmental vs. unrelated)|subject slope	0.1138
training attempt within session|subject slope	0.1508
session|subject slope	0.0502
frequency (Kucera-Francis)|subject slope	0.0189
length (letters)|subject slope	0.0387
days since last training session|subject slope	0.0139
trials since last trained|subject slope	0.0208
block type (segmental vs. unrelated)*training attempt within session|subject slope	0.0222
block type (segmental vs. unrelated)*session|subject slope	0.0296
training attempt within session*session|subject slope	0.0036
block type (segmental vs. unrelated)*training attempt within session*session|subject slope	0.0005

Table F3. Results of the analysis of the semantic model of the assessment data including all time points.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	0.92	0.29	3.15	.002
block type (semantic vs. unrelated)	-0.04	0.12	-0.34	.735
time (before vs. after training)	1.04	0.11	9.18	<.001
time (immediately post-training vs. follow-up)	-0.53	0.13	-4.15	<.001
frequency (Kucera-Francis)	0.01	0.12	0.11	.914
length (letters)	-0.21	0.15	-1.41	.157
block type (semantic vs. unrelated)*time (before vs. after training)	-0.01	0.10	-0.14	.887
block type (semantic vs. unrelated)*time (immediately post-training vs. follow-up)	-0.08	0.13	-0.60	.548

Random effects	Variance
item intercept	0.7406
time (before vs. after training)|item slope	0.4015
time (immediately post-training vs. follow-up)|item slope	0.3937
subject intercept	0.5028
block type (semantic vs. unrelated)|subject slope	0.0640
time (before vs. after training)|subject slope	0.0247
time (immediately post-training vs. follow-up)|subject slope	0.0285
frequency (Kucera-Francis)|subject slope	0.0181
length (letters)|subject slope	0.1022
block type (semantic vs. unrelated)*time (before vs. after training)|subject slope	0.0364
block type (semantic vs. unrelated)*time (immediately post-training vs. follow-up)|subject slope	0.0682

Table F4. Results of the analysis of the segmental model of the assessment data including all time points.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	0.95	0.29	3.27	.001
block type (segmental vs. unrelated)	0.20	0.10	2.04	.041
time (before vs. after training)	1.08	0.16	6.65	<.001
time (immediately post-training vs. follow-up)	-0.56	0.13	-4.20	<.001
frequency (Kucera-Francis)	0.07	0.09	0.87	.384
length (letters)	-0.21	0.15	-1.38	.167
block type (segmental vs. unrelated)*time (before vs. after training)	0.13	0.09	1.40	.160
block type (segmental vs. unrelated)*time (immediately post-training vs. follow-up)	-0.20	0.10	-2.15	.032

Random effects	Variance
item intercept	0.5891
time (before vs. after training)|item slope	0.4002
time (immediately post-training vs. follow-up)|item slope	0.1762
subject intercept	0.5148
block type (segmental vs. unrelated)|subject slope	0.0284
time (before vs. after training)|subject slope	0.1252
time (immediately post-training vs. follow-up)|subject slope	0.0622
frequency (Kucera-Francis)|subject slope	0.0020
length (letters)|subject slope	0.1275
block type (segmental vs. unrelated)*time (before vs. after training)|subject slope	0.0225
block type (segmental vs. unrelated)*time (immediately post-training vs. follow-up)|subject slope	0.0247

Table F5. Results of the analysis of the semantic model of the assessment data comparing performance before training to performance immediately after training.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	0.93	0.27	3.42	.001
block type (semantic vs. unrelated)	0.05	0.15	0.34	.731
time (pre-training vs. immediately post-training)	1.05	0.14	7.38	<.001
frequency (Kucera-Francis)	0.02	0.10	0.23	.817
length (letters)	-0.18	0.14	-1.30	.193
block type (semantic vs. unrelated)*time (pre-training vs. immediately post-training)	0.05	0.13	0.43	.665

Random effects	Variance
item intercept	0.9256
time (pre-training vs. immediately post-training)|item slope	0.5368
subject intercept	0.3862
block type (semantic vs. unrelated)|subject slope	0.1055
time (pre-training vs. immediately post-training)|subject slope	0.0439
frequency (Kucera-Francis)|subject slope	0.0070
length (letters)|subject slope	0.0841
block type (semantic vs. unrelated)*time (pre-training vs. immediately post-training)|subject slope	0.0624

Table F6. Results of the analysis of the segmental model of the assessment data comparing performance before training to performance immediately after training.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	1.04	0.27	3.83	<.001
block type (segmental vs. unrelated)	0.26	0.12	2.16	.031
time (pre-training vs. immediately post-training)	1.13	0.16	6.93	<.001
frequency (Kucera-Francis)	0.08	0.09	0.86	.393
length (letters)	-0.19	0.15	-1.28	.201
block type (segmental vs. unrelated)*time (pre-training vs. immediately post-training)	0.20	0.09	2.22	.026

Random effects	Variance
item intercept	0.6954
time (pre-training vs. immediately post-training)|item slope	0.4265
subject intercept	0.4094
block type (segmental vs. unrelated)|subject slope	0.0469
time (pre-training vs. immediately post-training)|subject slope	0.1004
frequency (Kucera-Francis)|subject slope	0.0005
length (letters)|subject slope	0.1132
block type (segmental vs. unrelated)*time (pre-training vs. immediately post-training)|subject slope	0.0082

Table F7. Results of the analysis of the semantic model of the assessment data comparing performance before training to performance at follow-up.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	0.37	0.28	1.36	.173
block type (semantic vs. unrelated)	-0.07	0.09	-0.79	.430
time (pre-training vs. follow-up)	0.51	0.07	7.30	<.001
frequency (Kucera-Francis)	-0.05	0.10	-0.47	.636
length (letters)	-0.21	0.14	-1.47	.143
block type (semantic vs. unrelated)*time (pre-training vs. follow-up)	-0.05	0.06	-0.79	.430

Random effects	Variance
item intercept	0.4661
time (pre-training vs. follow-up)|item slope	0.1018
subject intercept	0.4788
block type (semantic vs. unrelated)|subject slope	0.0233
time (pre-training vs. follow-up)|subject slope	0.0133
frequency (Kucera-Francis)|subject slope	0.0133
length (letters)|subject slope	0.0955
block type (semantic vs. unrelated)*time (pre-training vs. follow-up)|subject slope	0.0102

Table F8. Results of the analysis of the segmental model of the assessment data comparing performance before training to performance at follow-up.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	0.40	0.30	1.33	.183
block type (segmental vs. unrelated)	0.05	0.07	0.74	.458
time (pre-training vs. follow-up)	0.53	0.11	4.63	<.001
frequency (Kucera-Francis)	0.07	0.08	0.83	.409
length (letters)	-0.21	0.14	-1.48	.139
block type (segmental vs. unrelated)*time (pre-training vs. follow-up)	-0.01	0.07	-0.15	.882

Random effects	Variance
item intercept	0.3488
time (pre-training vs. follow-up)|item slope	0.1563
subject intercept	0.5925
block type (segmental vs. unrelated)|subject slope	0.0068
time (pre-training vs. follow-up)|subject slope	0.0676
frequency (Kucera-Francis)|subject slope	0.0014
length (letters)|subject slope	0.1049
block type (segmental vs. unrelated)*time (pre-training vs. follow-up)|subject slope	0.0132

Appendix G: Generalization Analysis from Study 2

In this analysis, effects of training words in blocks on generalization to untrained words were examined, evaluating whether there is support for the predictions regarding changes to distinctiveness as a result of blocked training. At a broader level, the analysis also allowed investigation of whether there is generalization at all as a result of this training protocol

Model structure. As in the previous analyses presented in this study, multilevel mixed effects models with random effects were constructed to examine group letter accuracy data on the generalization list before and after treatment. The full generalization list contained all items that were trained as well as untrained items from screening and baseline assessments. Four total generalization models were constructed. In half the models, both trained and untrained items were included so that effects of training in blocks on trained words could be compared to effects of training in blocks on untrained words. The other half consisted of models in which only untrained items were included to see if training other items in blocks led to improvement in related words as compared to words that were not related to the trained set. As in all previous analyses, separate models compared semantic vs. unrelated block type and segmental vs. unrelated block type.

Models included treatment group (untrained or trained), block type[25] (semantic [trained in a semantic block or semantically related to items trained in a semantic block],

[25] Note that block type refers not only to the training context for treated words, but also to whether the untreated generalization words are related to items trained in the semantic and segmental blocks. For example, if treatment words in a semantic block refer to fruit, an additional generalization word that refers to a fruit would be treated as a belonging to a semantic block. If treatment words in a segmental block all contain the letters OR, an additional generalization word that contains those letters would be treated as a belonging to a segmental block.

segmental [trained in a segmental block or segmentally related to items trained in a segmental block], or unrelated [trained in an unrelated context or unrelated to items trained in semantic and segmental blocks]), and time point (after or before training), the two- and three-way interaction between them, and the control variables of word frequency and word length as fixed effects. A full random structure was implemented in each model, with random intercepts for subjects and items and a full random slope structure matching the fixed effect structure, excluding random slopes over items for block type, frequency, and length since each item has only one value for these variables.

The continuous variables of frequency and length were centered and scaled. The categorical variable of time point was coded as before training=-1, after training=1. The categorical variable of treatment group was coded as trained=-1, untrained=1 for the models including both trained and untrained items. Separate models compared semantic=1 vs. unrelated=-1 block types, excluding the segmental block type, or compared segmental=1 vs. unrelated=-1 block types, excluding the semantic block type.

Explanation of included variables. In the generalization analyses, the effect of time point evaluates whether participants produce items more accurately after training, even if those items themselves were not trained.

The effect of treatment group evaluates whether participants produce trained items more accurately than untrained items. As a caveat, recall that the simple main effect includes responses that occurred before training, when performance for the items that were to be trained may or may not differ from the items that were never trained.

The interaction between time point and treatment group evaluates whether trained words show a different change in accuracy from before treatment to after treatment as compared to untrained items. A significant negative interaction would be expected if learning is effective but does not generalize to untrained items, or if the generalization effect is smaller than the treatment effect. The interaction would not be significant if training is ineffective and does not increase performance of trained words nor does it lead to generalization, or if treatment is effective and there is generalization to untrained items. To distinguish between these possibilities, one can look at the effect of time point in the models including only untrained items. If training is effective and leads to generalization, the effect of time point would be significant and positive, whereas time point would not be significant if there is not generalization to untrained items.

The effect of block type evaluates whether items trained in a particular context and related untrained items are produced more accurately than items in a comparison context.

The critical effect for examining whether the e-ILM's predicted distinctiveness changes due to training in blocks result in differences in generalization to semantically and segmentally related words is the interaction of block type and time. This interaction evaluates whether there is more or less improvement from before to after treatment for semantically or segmentally related items relative to comparison unrelated items. In the models including trained and untrained items, a significant interaction could be driven by improvement in trained items from a particular context. Therefore, it is crucial to look at the interaction of block type and time in the models that include only untrained items. Here, a positive interaction would suggest that items related to those in trained in the particular blocked context improve more than the comparison unrelated items, indicating successful generalization. A negative interaction would suggest that items related to those in trained in the particular blocked context improve less than the comparison unrelated items, indicating that training in blocks was detrimental to related items and did not lead to successful generalization. If the predictions of the extended incremental learning model hold for this group, reduced generalization is expected for items that are semantically related to those trained in semantic blocks. The predictions are less clear regarding the effects on generalization of training in segmental blocks vs. unrelated contexts, so a positive interaction, negative interaction, or null interaction may be found.

The two-way interaction of block type and treatment evaluates whether items trained in a particular context differ from related items that were not trained. This only exists in the model including both trained and untrained items, as does the three-way interaction of block type, time point, and treatment group. This three-way interaction evaluates whether items trained in a particular context improve more or less than related items that were not trained over time.

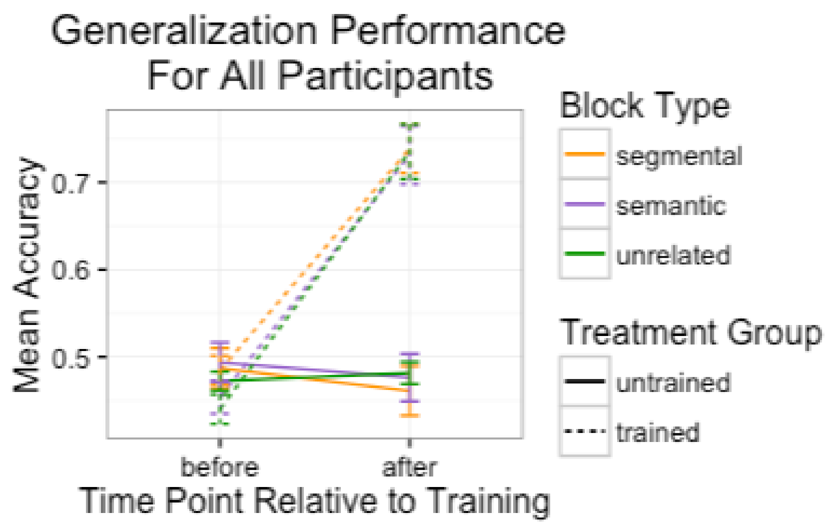
Frequency and length of the words being trained were also included to control for psycholinguistic factors that are known to affect the performance of these individuals.

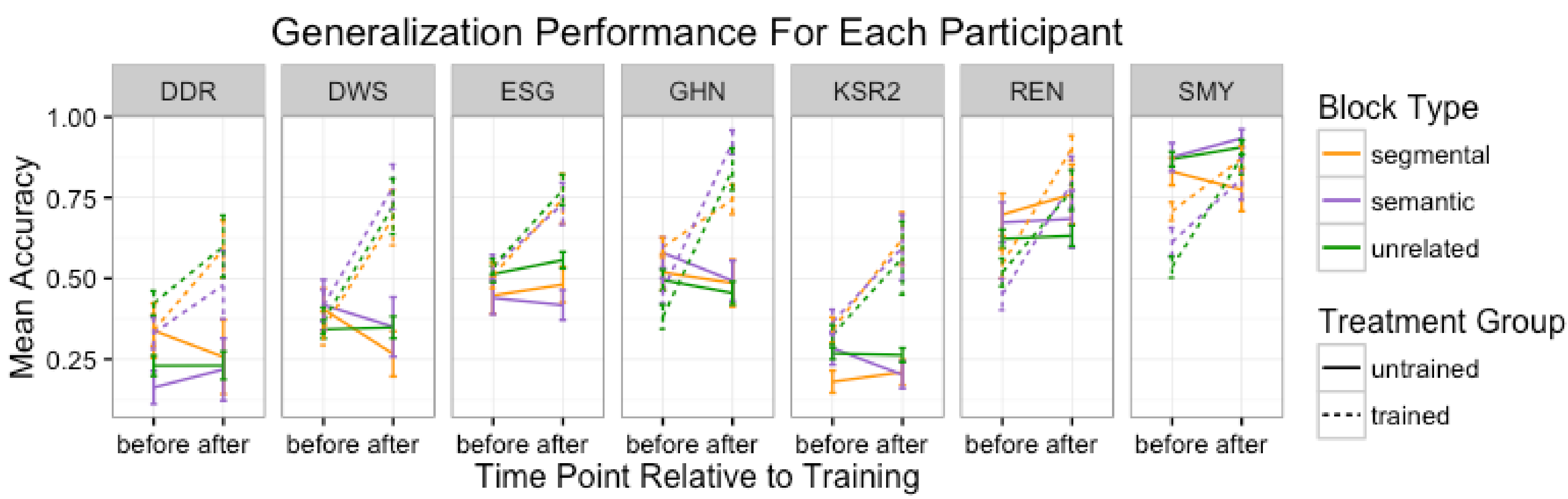
Results of generalization analysis. Figure G1 show performance on the generalization task administered before and after treatment. Tables G1-G4 (at the end of this appendix) report the results of the analyses of this task.

Figure G1. Results of the generalization task in the dysgraphia treatment study.

The top panel shows mean accuracy, measured as proportion of letters spelled correctly in each word, for all seven participants during the generalization task in which they named a large set of items before and after treatment. Dotted lines represent items that were trained; solid lines represent untrained items. Different colored lines represent the segmental, semantic, and unrelated contexts. Trained items were practiced in segmental, semantic, or unrelated blocks. Untrained segmental items shared letters with the items trained in segmental blocks; untrained semantic blocks were from the same category as items trained in semantic blocks; and untrained unrelated items did not fit either of the other characterizations. Error bars represent one standard error of the mean, corrected for repeated measures.

The bottom panel shows mean accuracy on the generalization task for each of the seven individual participants. Error bars represent one standard error of the mean.


Generalization Performance
For All Participants
Mean Accuracy
0.7
0.6
0.5
before
after
Time Point Relative to Training
Block Type
segmental
semantic
unrelated
Treatment Group
untrained
trained


Generalization Performance For Each Participant
DDR
DWS
ESG
GHN
KSR2
REN
SMY
Mean Accuracy
1.00
0.75
0.50
0.25
before after
Time Point Relative to Training
Block Type
segmental
semantic
unrelated
Treatment Group
untrained
trained

The critical effect for evaluating whether training in blocks impacted generalization to related words was the interaction of block type and time. If training in semantic blocks increases distinctiveness as predicted in the extended incremental learning model, reduced generalization should be observed for untrained items that are semantically related to items trained in semantic blocks relative to unrelated items. If training in segmental blocks reduces distinctiveness as predicted in the extended incremental learning model, increased generalization may be observed for untrained items that are segmentally related to items trained in segmental blocks relative to unrelated items because of the strengthening of connections between shared segments and lexical nodes, reduced generalization may be observed because of the weakening of connections between distinctive segments and lexical nodes, or a null effect may be observed because the strengthening and weakening of connections cancel each other out. At the group level, there were not significant interactions of block type and time in the models that included both trained and untrained items or in the models looking only at untrained items. For untrained words, being semantically or segmentally related to the training blocks used in treatment did not result in better or worse performance than being unrelated to the training blocks; there was neither increased nor decreased generalization. However, as in the previously reported analysis supporting the idea that semantic blocking during training is a desirable difficulty, it could be that there are effects at the individual level, which will be investigated below. The effect observed here does not support the predictions of the extended incremental learning model regarding the effects of training in semantic blocks vs. unrelated contexts, although it is consistent with a balance between strengthening of shared features and weakening of distinctive features that leads to no effect on generalization of training in segmental blocks vs. unrelated contexts.

The generalization analyses did show that there were differences in trained and untrained items as a result of treatment. In the models that included both trained and untrained items, there was a consistent main effect of time point such that responses were more accurate after training than before training (z=6.54, p<.001 for the semantic model; z=7.27, p<.001 for the segmental model). There was also a consistent main effect of treatment group such that responses to untrained words were less accurate than responses to trained words (z=-1.73, p=.083 for the semantic model; z=-3.04, p=.002 for the segmental model). The two-way interaction between time point and treatment group was also consistently significant, indicating that there was greater improvement for trained words than for untrained words over the course of the study (z=-5.30, p<.001 for the semantic model; z=-7.70, p<.001 for the segmental model). Together, these results show that while there was overall improvement from the beginning to the end of the study, trained items improved more than untrained items.

In terms of the effects of block type on trained vs. untrained words, there was a marginally significant two-way interaction between block type and treatment group in the segmental model: participants showed a trend for better performance for words that were treated in segmental blocks as compared to unrelated blocks than for segmentally related words that were not treated vs. unrelated words that were not treated (z=-1.70, p=.089). That there was no significant interaction of this effect with time suggests that participants had a larger accuracy advantage for words that were to be trained in segmental blocks as opposed to unrelated blocks than they did for the segmentally related words that were not to be trained relative to unrelated words that were not to be trained even before training commenced. This is consistent with the simple main effects of segmental vs. unrelated block type found in the previous training and assessment analyses: participants were consistently better with the words in the segmental training blocks than with the words in the unrelated training blocks. There was not a corresponding significant two-way interaction between block type and treatment group in the semantic model, not were there no significant main effects of block type, two-way interactions of block type with time point, or three-way interactions of block type, time point, and treatment group in either the semantic or the segmental models that included both trained and untrained items.

When considering only untrained words, there was no evidence to suggest that these items benefitted from treatment. There were not significant main effects of time point, indicating that untrained words did not significantly improve after treatment. There were not significant effects of block type, indicating that words related to training items were no more or less accurately produced than unrelated items. Above, it was already discussed that there were no significant interactions of block type and time point. Regardless of their relationship to trained items, untrained items did not improve after treatment; there was no evidence of generalization to untrained items for the group as a whole.

There were effects of the psycholinguistic variable of length. Participants responded more accurately in shorter words. This effect was significant in all models (z=-3.29, p=.001 for the semantic model including trained and untrained words; z=-3.09, p=.002 for the segmental model including trained and untrained words; z=-3.63, p<.001

for the semantic model including only untrained words; z=-3.45, p=.001 for the segmental model including only untrained words).There were not significant effects of frequency.

Discussion of generalization analysis. Overall, these results indicate that there was not generalization of learning from trained items to untrained items, regardless of the relationship between the untrained items and training contexts. The improvement over time seen in the models that included trained items was likely driven by improvement of the trained items. For the participant group as a whole, the treatment protocol appears to have had item-specific effects, rather than improvement that spread to other related or unrelated items. The effects found are not in line with the extended incremental learning model, which predicted that there would be reduced generalization for untrained items that were semantically related vs. unrelated to items trained in semantic blocks. However, with regards to the more exploratory analysis of the effects of segmental blocking on generalization, the results suggest that the reduced distinctiveness caused by training in segmentally related vs. unrelated blocks does not affect generalization, possibly because the weakening of connections between distinctive segments and lexical nodes that might otherwise lead to reduced generalization is canceled out by the strengthening of connections between shared segments and lexical nodes that might otherwise lead to increased generalization. Note that this finding is also consistent with failure of the extended incremental learning model: if training in segmental vs. unrelated blocks does not change distinctiveness, the null effect seen here would be expected.

Table G1. Results of the analysis of the semantic model of the generalization data including both trained and untrained items.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	0.22	0.37	0.60	.547
time point (after vs. before training)	0.34	0.05	6.54	<.001
treatment group (untrained vs. trained)	-0.24	0.14	-1.73	.083
block type (semantic vs. unrelated)	0.02	0.06	0.27	.784
frequency (Kucera-Francis)	0.06	0.05	1.17	.242
length (letters)	-0.23	0.07	-3.29	.001
time point (after vs. before training)*treatment group (untrained vs. trained)	-0.32	0.06	-5.30	<.001
time point (after vs. before training)*block type (semantic vs. unrelated)	-0.02	0.04	-0.54	.588
treatment group (untrained vs. trained)*block type (semantic vs. unrelated)	-0.02	0.05	-0.40	.693
time point (after vs. before training)*treatment group (untrained vs. trained)*block type (semantic vs. unrelated)	-0.04	0.04	-0.92	.360

Random effects	Variance
item intercept	0.5323
time point (after vs. before training)|item slope	0.0346
subject intercept	0.9365
time point (after vs. before training)|subject slope	0.0107
treatment group (untrained vs. trained)|subject slope	0.1247
block type (semantic vs. unrelated)|subject slope	0.0100
frequency (Kucera-Francis)|subject slope	0.0032
length (letters)|subject slope	0.0166
time point (after vs. before training)*treatment group (untrained vs. trained)|subject slope	0.0180
time point (after vs. before training)*block type (semantic vs. unrelated)|subject slope	0.0024
treatment group (untrained vs. trained)*block type (semantic vs. unrelated)|subject slope	0.0086
time point (after vs. before training)*treatment group (untrained vs. trained)*block type (semantic vs. unrelated)|subject slope	0.0041

Table G2. Results of the analysis of the segmental model of the generalization data including both trained and untrained items.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	0.20	0.37	0.55	.582
time point (after vs. before training)	0.34	0.05	7.27	<.001
treatment group (untrained vs. trained)	-0.29	0.10	-3.04	.002
block type (segmental vs. unrelated)	0.03	0.06	0.46	.643
frequency (Kucera-Francis)	0.06	0.06	1.06	.287
length (letters)	-0.21	0.07	-3.09	.002
time point (after vs. before training)*treatment group (untrained vs. trained)	-0.31	0.04	-7.70	<.001
time point (after vs. before training)*block type (segmental vs. unrelated)	-0.02	0.05	-0.49	.625
treatment group (untrained vs. trained)*block type (segmental vs. unrelated)	-0.09	0.05	-1.70	.089
time point (after vs. before training)*treatment group (untrained vs. trained)*block type (segmental vs. unrelated)	-0.02	0.04	-0.47	.637

Random effects	Variance
item intercept	0.5178
time point (after vs. before training)|item slope	0.0473
subject intercept	0.9156
time point (after vs. before training)|subject slope	0.0074
treatment group (untrained vs. trained)|subject slope	0.0533
block type (segmental vs. unrelated)|subject slope	0.0110
frequency (Kucera-Francis)|subject slope	0.0058
length (letters)|subject slope	0.0184
time point (after vs. before training)*treatment group (untrained vs. trained)|subject slope	0.0042
time point (after vs. before training)*block type (segmental vs. unrelated)|subject slope	0.0105
treatment group (untrained vs. trained)*block type (segmental vs. unrelated)|subject slope	0.0077
time point (after vs. before training)*treatment group (untrained vs. trained)*block type (segmental vs. unrelated)|subject slope	0.0052

Table G3. Results of the analysis of the semantic model of the generalization data including only untrained items.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	-0.03	0.52	-0.05	.959
time point (after vs. before training)	0.03	0.05	0.47	.640
block type (semantic vs. unrelated)	-0.03	0.08	-0.33	.744
frequency (Kucera-Francis)	0.05	0.06	0.81	.419
length (letters)	-0.29	0.08	-3.63	<.001
time point (after vs. before training)*block type (semantic vs. unrelated)	-0.05	0.04	-1.20	.232

Random effects	Variance
item intercept	0.6707
time point (after vs. before training)|item slope	0.0495
subject intercept	1.8537
time point (after vs. before training)|subject slope	0.0077
block type (semantic vs. unrelated)|subject slope	0.0249
frequency (Kucera-Francis)|subject slope	0.0055
length (letters)|subject slope	0.0176
time point (after vs. before training)*block type (semantic vs. unrelated)|subject slope	0.0004

Table G4. Results of the analysis of the segmental model of the generalization data including only untrained items.

Fixed effects	Coefficient	SE	*z*	*p*
Intercept	-0.12	0.47	-0.26	.797
time point (after vs. before training)	0.04	0.05	0.72	.475
block type (segmental vs. unrelated)	-0.11	0.07	-1.42	.156
frequency (Kucera-Francis)	0.06	0.07	0.94	.346
length (letters)	-0.29	0.08	-3.45	.001
time point (after vs. before training)*block type (segmental vs. unrelated)	-0.04	0.05	-0.85	.398

Random effects	Variance
item intercept	0.7153
time point (after vs. before training)|item slope	0.0557
subject intercept	1.5399
time point (after vs. before training)|subject slope	0.0058
block type (segmental vs. unrelated)|subject slope	0.0186
frequency (Kucera-Francis)|subject slope	0.0086
length (letters)|subject slope	0.0217
time point (after vs. before training)*block type (segmental vs. unrelated)|subject slope	0.0053

REFERENCES

Alario, F.-X., De Cara, B., & Ziegler, J. C. (2007). Automatic activation of phonology in silent reading is parallel: evidence from beginning and skilled readers. *Journal of Experimental Child Psychology*, *97*, 205–19.

Altarriba, J., & Mathis, K. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language*, *36*, 550–568.

Anderson, N. D., & Craik, F. I. M. (2006). The mnemonic mechanisms of errorless learning. *Neuropsychologia*, *44*, 2806–2813.

Baddeley, A. D., & Wilson, B. a. (1994). When implicit learning fails: amnesia and the problem of error elimination. *Neuropsychologia*, *32*, 53–68.

Ball, A. L., de Riesthal, M., Breeding, V. E., & Mendoza, D. E. (2011). Modified act and cart in severe aphasia. *Aphasiology*, *25*, 836–848.

Basso, A., Marangolo, P., Piras, F., & Galluzzi, C. (2001). Acquisition of new "words" in normal subjects: a suggestion for the treatment of anomia. *Brain and Language*, *77*, 45–59.

Beeson, P. M. (1999). Treating acquired writing impairment : strengthening graphemic representations. *Aphasiology*, *13*, 767–785.

Beeson, P. M. (2004). Remediation of written language. *Topics in Stroke Rehabilitation*, *11*, 37–48.

Beeson, P. M., & Egnor, H. (2006). Combining treatment for written and spoken naming. *Journal of the International Neuropsychological Society : JINS*, *12*, 816–27.

Beeson, P. M., Hirsch, F. M., & Rewega, M. a. (2002). Successful single-word writing treatment: experimental analyses of four cases. *Aphasiology*, *16*, 473–491.

Beeson, P. M., & Rapcsak, S. Z. (2012). Neuropsychological assessment and rehabilitation of writing disorders. In J. Gurd, U. Kischka, & J. Marshall (Eds.), *The Handbook of Clinical Neuropsychology* (pp. 725–775). Oxford: Oxford University Press.

Beeson, P. M., Rising, K., & Volk, J. (2003). Writing treatment for severe aphasia: who benefits? *Journal of Speech, Language, and Hearing Research : JSLHR*, *46*, 1038–60.

Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review*, *15*, 357–363.

Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: implications for models of lexical-semantic encoding. *Journal of Memory and Language*, *69*, 228–256.

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: evidence for a lexical-semantic competition hypothesis. *Cognition*, *96*, 45–54.

Belke, E., & Meyer, A. S. (2007). Single and multiple object naming in healthy ageing. *Language and Cognitive Processes*, *22*, 1178–1211.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology Section A*, *58*, 667–692.

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: evidence from blocked and continuous manipulations of semantic context. *Quarterly Journal of Experimental Psychology (2006)*, *66*, 2135–60.

Bier, N., Van Der Linden, M., Gagnon, L., Desrosiers, J., Adam, S., Louveaux, S., & Saint-Mleux, J. (2008). Face-name association learning in early alzheimer's disease: a comparison of learning methods and their underlying mechanisms. *Neuropsychological Rehabilitation*, *18*, 343–371.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*, 2767–96.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* (pp. 56–64). New York: Worth Publishers.

Bjork, E. L., deWinstanley, P. A., & Storm, B. C. (2007). Learning how to learn: can experiencing the outcome of different encoding strategies enhance subsequent encoding? *Psychonomic Bulletin & Review*, *14*, 207–11.

Bjork, R. A. (1975). Retrieval as a memory modifier: an interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A. (2013). Desirable difficulties perspective on learning. In H. Pashler (Ed.), *Encyclopedia of the Mind*. Thousand Oaks: Sage Reference. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Desirable+Difficulties+Perspective+on+Learning#2\nhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Desirable+difficulties+perspective+on+learning#2

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes* (pp. 35–67). Hillsdale, NJ: Erlbaum.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology (London, England : 1953)*, *93*, 89–114.

Bonin, P., & Fayol, M. (2000). Writing words from pictures: what representations are activated, and when? *Memory & Cognition*, *28*, 677–89.

Bonin, P., Fayol, M., & Peereman, R. (1998). Masked form priming in writing words from pictures: evidence for direct retrieval of orthographic codes. *Acta Psychologica*, *99*, 311–328.

Bonin, P., Méot, A., Lagarrigue, A., & Roux, S. (2015). Written object naming, spelling to dictation, and immediate copying: different tasks, different pathways? *Quarterly Journal of Experimental Psychology*, *68*, 1268–1294.

Bousfield, W. A. (1953). The occurence of clustering in the recall of randomly arranged associates. *The Journal of General Psychology*, *49*, 229–240.

Bower, G., Clark, M., Lesgold, A., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343.

Bradlow, A. R. (2008). Training non-native language sound patterns: lessons from training japanese adults on the english /r/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 287–308). John Benjamins Publishing Company.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training japanese listeners to identify engish /r/ and /l/: long-term retention of learning in perception and production. *Perceptual Psychophysics*, *61*, 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training japanese listeners to identify english /r/ and /l/: iv. some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*, 2299–2310.

Breining, B. L., Lala, T., Cuitiño, M. M., Manes, F., Peristeri, E., Tsapkini, K., … Hillis, A. E. (2015). A brief assessment of object semantics in primary progressive aphasia. *Aphasiology*, *29*, 488–505.

Breining, B. L., Nozari, N., & Rapp, B. C. (2016). Does segmental overlap help or hurt? evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, *23*, 500–506.

Breining, B. L., & Rapp, B. C. (submitted). Shared principles of word production in writing and speaking.

Brown, A. S. (1981). Inhibition in cued retrieval. *Journal of Experimental Psychology: Human Learning & Memory*, *7*, 204–215.

Buchwald, A., & Rapp, B. C. (2009). Distinctions between orthographic long-term memory and working memory. *Cognitive Neuropsychology*, *26*, 724–751.

Capitani, E., Laiacona, M., Mahon, B. Z., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? a critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*, 213–261.

Cappa, S. F., Perani, D., Schnur, T. T., Tettamanti, M., & Fazio, F. (1998). The effects of semantic category and knowledge type on lexical-semantic access: a pet study. *NeuroImage*, *8*, 350–9.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, *14*, 177–208.

Caramazza, A., & Hillis, A. E. (1990). Where do semantic errors come from? *Cortex*, *26*, 95–122.

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, *7*, 354–361.

Caramazza, A., & Mahon, B. Z. (2006). The organisation of conceptual knowledge in the brain: the future's past and some future directions. *Cognitive Neuropsychology*, *23*, 13–38.

Caramazza, A., Miceli, G., Villa, G., & Romani, C. (1987). The role of the graphemic buffer in spelling: evidence from a case of acquired dysgraphia. *Cognition*, *26*, 59–85.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1–34.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–80.

Channell, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal*, *35*, 115–122.

Chao, L. L., Haxby, J. V, & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*, 913–9.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, *12*, 478–84.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex (New York, N.Y. : 1991)*, *12*, 545–51.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*, 417–30.

Clare, L., & Jones, R. S. P. (2008). Errorless learning in the rehabilitation of memory impairment: a critical review. *Neuropsychology Review*, *18*, 1–23.

Cole, M., Frankel, F., & Sharp, D. (1971). Development of free recall learning in children. *Developmental Psychology*, *4*, 109–123.

Coltheart, M. (1981). The mrc psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505.

Coltheart, M., Davelaar, E., Jonassen, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance, Vol. 6* (Lawrence E., pp.

535–556). Hillsdale, NJ.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–56.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology*, *23*, 1311–1337.

Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., & Salmelin, R. (2004). Learning new names for new objects: cortical effects as measured by magnetoencephalography. *Brain and Language*, *89*, 617–622.

Costa, A., & Caramazza, A. (2002). The production of noun phrases in english and spanish: implications for the scope of phonological encoding in speech production. *Journal of Memory and Language*, *46*, 178–198.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: implications for theories of semantic memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 643–658.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: segment type and order information in early lexical entries. *Journal of Memory and Language*, *54*, 1–19.

Crowther, J. E., & Martin, R. C. (2014). Lexical selection in the semantically blocked cyclic naming task: the role of cognitive control and learning. *Frontiers in Human*

Neuroscience, *8*, 1–20.

Crutch, S. J., Connell, S., & Warrington, E. K. (2009). The different representational frameworks underpinning abstract and concrete knowledge: evidence from odd-one-out judgements. *Quarterly Journal of Experimental Psychology (2006)*, *62*, 1377–88, 1388–90.

Crutch, S. J., Ridha, B. H., & Warrington, E. K. (2006). The different frameworks underlying abstract and concrete knowledge: evidence from a bilingual patient with a semantic refractory access dysphasia. *Neurocase : Case Studies in Neuropsychology, Neuropsychiatry, and Behavioural Neurology*, *12*, 151–63.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain : A Journal of Neurology*, *128*, 615–27.

Crutch, S. J., & Warrington, E. K. (2007). Semantic priming in deep-phonological dyslexia: contrasting effects of association and similarity upon abstract and concrete word reading. *Cognitive Neuropsychology*, *24*, 583–602.

Crutch, S. J., & Warrington, E. K. (2010). The differential dependence of abstract and concrete words upon associative and similarity-based information: complementary semantic interference and facilitation effects. *Cognitive Neuropsychology*, *27*, 46–71.

D'Agostino, P. R. (1969). The blocked-random effect in recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *8*, 815–820.

Damasio, H., Tranel, D., Grabowski, T. J., Adolphs, R., & Damasio, A. R. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*, 179–229.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 416–431.

Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*, 1372–84.

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language*, *49*, 119–132.

Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, *24*, 685–712.

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*, B77–86.

Dell, G. S., Nozari, N., & Oppenheim, G. M. (2014). Word production: behavioral and computational considerations. In M. Goldrick & M. Miozzo (Eds.), *The Oxford Handbook of Language Production*. New York: Oxford University Press.

Devlin, J. T., Moore, C. J., Mummery, C. J., Gorno-Tempini, M. L., Phillips, J. a, Noppeney, U., … Price, C. J. (2002). Anatomic constraints on cognitive theories of category specificity. *NeuroImage*, *15*, 675–85.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J., & Tyler, L. K. (2002). Is there an anatomical basis for category-specificity? semantic memory studies in pet and fmri. *Neuropsychologia*, *40*, 54–75.

Dunn, J., & Clare, L. (2007). Learning face-name associations in early-stage dementia: comparing the effects of errorless learning and effortful processing. *Neuropsychological Rehabilitation*, *17*, 735–754.

Ebbinghaus, H. (1913). *Memory: a contribution to experimental psychology*. (H. A. Ruger, C. E. Bussenius, & Translators, Eds.). New York: Teachers College, Columbia University. Retrieved from http://psy.ed.asu.edu/~classics/Ebbinghaus/index.htm

Ellis, A. W., & Young, A. W. (1988). *Human cognitive neuropsychology*. (A. W. Ellis & A. W. Young, Eds.)*Human Cognitive Neuropsychology*. Hove, UK: Erlbaum.

Erten, İ. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, *36*, 407–422.

Evans, J. J., Wilson, B. A., Schuri, U., Andrade, J., Baddeley, A. D., Bruna, O., … Taussik, I. (2000). A comparison of “errorless” and “trial-and-error” learning methods for teaching individuals with acquired memory deficits. *Neuropsychological Rehabilitation*, *10*, 67–101.

Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: a review of theory and practice. *Neuropsychological Rehabilitation*, *13*, 337–63.

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2005a). Treatment of anomia using errorless versus errorful learning: are frontal executive skills and feedback important? *International Journal of Language & Communication Disorders*, *40*, 505–23.

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2005b). Further explorations and an overview of errorless and errorful therapy for aphasic word-finding difficulties: the number of naming attempts during therapy affects outcome. *Aphasiology*, *19*, 597–614.

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation*, *16*, 129–154.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, *24*, 369–383.

Freed, D. B., Marshall, R. C., & Phillips, D. S. (1998). Comparison of semantically and phonemically based training procedures in an overlearned naming task. *Perceptual and Motor Skills*, *87*, 795–800.

Gerlach, C. (2007). A review of functional imaging studies on category specificity. *Journal of Cognitive Neuroscience*, *19*, 296–314.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Distinct and common cortical activations for multimodal semantic categories. *Cognitive, Affective & Behavioral Neuroscience*, *6*, 214–22.

Goldrick, M., Folk, J. R., & Rapp, B. C. (2010). Mrs. malaprop's neighborhood: using word errors to reveal neighborhood structure. *Journal of Memory and Language*, *62*, 113–134.

Goldrick, M., & Rapp, B. C. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*, 219–260.

Gollin, E. S., & Sharps, M. J. (1988). Facilitation of free recall by categorical blocking depends on stimulus type. *Memory & Cognition*, *16*, 539–44.

Goodman, R. A., & Caramazza, A. (1985). *The johns hopkins university dysgraphia battery*. Baltimore: Johns Hopkins University.

Grönholm-Nyman, P., Rinne, J. O., & Laine, M. (2010). Learning and forgetting new names and objects in mci and ad. *Neuropsychologia*, *48*, 1079–1088.

Grönholm, P., Rinne, J. O., Vorobyev, V., & Laine, M. (2005). Naming of newly learned objects: a pet activation study. *Cognitive Brain Research*, *25*, 359–371.

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J. A., & Gee, J. (2002). The neural basis for category-specific knowledge: an fmri study. *NeuroImage*, *15*, 936–48.

Gupta, P., Martin, N., Abbs, B., Schwartz, M. F., & Lipinski, J. (2006). New word learning in aphasic patients: dissociating phonological and semantic components. *Brain and Language*, *99*, 8–9.

Hashemi, M. R., & Gowdaseaei, F. (2005). An attribute-treatment interaction study: lexical-set versus semantically-unrelated vocabulary instruction. *Regional Language Centre Journal*, *36*, 341–361.

Haslam, C., Moss, Z., & Hodder, K. (2010). Are two methods better than one? evaluating the effectiveness of combining errorless learning with vanishing cues. *Journal of Clinical and Experimental Neuropsychology*, *32*, 973–985.

Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, *2*, 170–175.

Hillis, A. E., & Caramazza, A. (1991). Mechanisms for accessing lexical representations for output: evidence from a category-specific semantic deficit. *Brain and Language*, *40*, 106–144.

Hillis, A. E., & Tippett, D. C. (2014). Stroke recovery : surprising influences and residual consequences. *Advances in Medicine*, *2014*, 1–10.

Hodgson, C., Schwartz, M. F., Schnur, T. T., & Brecher, A. (2005). Facilitation and interference in phonological blocked-cyclic naming. *Brain and Language*, *95*, 46–

47.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 562–567.

Horst, J. S. (2009). *Novel object & unusual name (noun) database*. Retrieved from http://www.sussex.ac.uk/wordlab/noun\npapers3://publication/uuid/0D203783-188B-4232-A478-04B6C8D76AAA

Hoshino, Y. (2010). The categorical facilitation effects on l2 vocabulary learning in a classroom setting. *RELC Journal*, *41*, 301–312.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*, 464–82.

Hultén, A., Vihla, M., Laine, M., & Salmelin, R. (2009). Accessing newly learned names and meanings in the native language. *Human Brain Mapping*, *30*, 976–989.

Humphreys, G. W., Evett, L., & Taylor, D. (1982). Automatic phonological priming in visual word recognition. *Memory & Cognition*, *10*, 576–590.

Hunkin, N. M., Squires, E. J., Parkin, A. J., & Tidy, J. a. (1998). Are the benefits of errorless learning dependent on implicit memory? *Neuropsychologia*, *36*, 25–36.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: a comparison of methods for teaching english /r/-/l/ to japanese adults. *The Journal of the Acoustical Society of America*, *118*, 3267–3278.

James, T. W., & Gauthier, I. (2004). Brain areas engaged during visual judgments by involuntary access to novel semantic information. *Vision Research*, *44*, 429–439.

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). *Retrieval-based learning: an episodic context account*. *The Psychology of Learning and Motivation* (Vol. 61). doi:10.1016/B978-0-12-800283-4.00007-1

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*, 704–19.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessments of language processing in aphasia*. East Sussex: Psychology Press.

Kelly, H., & Armstrong, L. (2009). New word learning in people with aphasia. *Aphasiology*, *23*, 1398–1417.

Kohnen, S., Nickels, L., Coltheart, M., & Brunsdon, R. (2008). Predicting generalization in the training of irregular-word spelling: treating lexical spelling deficits in a child. *Cognitive Neuropsychology*, *25*, 343–75.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219–24.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: is spacing the “enemy of induction”? *Psychological Science*, *19*, 585–592.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 989–98.

Kroll, J. F., & Stewart, E. (1994). Categorical interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Laiacona, M., Barbarotto, R., & Capitani, E. (2006). Human evolution and the brain representation of semantic knowledge: is there a role for sex differences? *Evolution and Human Behavior*, *27*, 158–168.

Laine, M., & Martin, N. (1996). Lexical retrieval deficit in picture naming: implications for word production models. *Brain and Language*, *53*, 283–314.

Laine, M., & Salmelin, R. (2010). Neurocognition of new word learning in the native tongue : lessons from the ancient farming equipment paradigm. *Language Learning*, *60*, 25–44.

Lambon Ralph, M. A., Snell, C., Fillingham, J. K., Conroy, P., & Sage, K. (2010). Predicting the outcome of anomia therapy for people with aphasia post cva: both language and cognitive status are key predictors. *Neuropsychological Rehabilitation*, *20*, 289–305.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training japanese listeners to identify english /r/ and /l/. ii: the role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training japanese listeners to identify english /r/ and /l/. iii. long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*, 2076–2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training japanese listeners to identify english /r/ and /l/: a first report. *The Journal of the Acoustical Society of America*, *89*, 874–886.

Lotto, L., & de Groot, a. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: a cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51.

Martin, A., & Caramazza, A. (2003). Neuropsychological and neuroimaging perspectives on conceptual knowledge: an introduction. *Cognitive Neuropsychology*, *20*, 195–212.

Martin, N., & Laine, M. (2000). Effects of contextual priming on impaired word retrieval. *Aphasiology*, *14*, 53–70.

Mayberry, E. J., Sage, K., Ehsan, S., & Lambon Ralph, M. A. (2011). Relearning in semantic dementia reflects contributions from both medial temporal lobe episodic and degraded neocortical semantic systems: evidence in support of the complementary learning systems theory. *Neuropsychologia*, *49*, 3591–8.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to japanese adults: tests of a hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective & Behavioral Neuroscience*, *2*, 89–108.

McCarthy, R. a., & Warrington, E. K. (1994). Disorders of semantic memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological*

Sciences, *346*, 89–96.

McClelland, J. L., Thomas, a G., McCandliss, B. D., & Fiez, J. A. (1999). Understanding failures of learning: hebbian learning, competition for representational space, and some preliminary experimental data. *Progress in Brain Research*, *121*, 75–80.

McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork* (pp. 175–198). New York: Taylor & Francis.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, *14*, 200–206.

McGeoch, J. A., & McDonald, W. T. (1931). Meaningful relation and retroactive inhibition. *The American Journal of Psychology*, *43*, 579–588.

McKissock, S., & Ward, J. (2007). Do errors matter? errorless and errorful learning in anomic picture naming. *Neuropsychological Rehabilitation*, *17*, 355–73.

McRae, K., Cree, G. S., Westmacott, R., & de Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology = Revue Canadienne de Psychologie Experimentale*, *53*, 360–373.

Metcalfe, J. (2011). Desirable difficulties and studying in the region of proximal learning. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork* (pp. 259–276). New York: Taylor & Francis.

Metzler-Baddeley, C., & Snowden, J. S. (2005). Brief report: errorless versus errorful learning as a memory rehabilitation approach in alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *27*, 1070–1079.

Meyer, A. S. (1990). The time course of phonological encoding in language production: the encoding of successive syllables of a word. *Journal of Memory and Language*, *29*, 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: phonological encoding inside a syllable. *Journal of Memory and Language*, *30*, 69–89.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 1146–1160.

Miceli, G., & Capasso, R. (1997). Semantic errors as neuropsychological evidence for the independence and the interaction of orthographic and phonological word forms. *Language and Cognitive Processes*, *12*, 733–764.

Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: a critical review. *Neuropsychological Rehabilitation*, *22*, 138–168.

Middleton, E. L., Schwartz, M. F., Rawson, K. A., & Garvey, K. (2015). Test-enhanced learning versus errorless learning in aphasia rehabilitation: testing competing psychological principles. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*, 1253–1261.

Mirman, D. (2014). *Growth curve analysis and visualization using r*. Boca Raton, FL: Chapman & Hall/CRC Press.

Moely, B., & Shapiro, S. (1971). Free recall and clustering at four age levels: effects of learning to learn and presentation method. *Developmental Psychology*, *4*, 490.

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain : A Journal of Neurology*, *122 (Pt 5*, 943–62.

Mulatti, C., Peressotti, F., Job, R., Saunders, S., & Coltheart, M. (2012). Reading aloud: the cumulative lexical interference effect. *Psychonomic Bulletin & Review*, *19*, 662–7.

Mummery, C. J., Patterson, K., Hodges, J. R., & Wise, R. J. . (1996). Generating "tiger" as an animal name or a word beginning with t: differences in brain activation. *Proceedings. Biological Sciences / The Royal Society*, *263*, 989–95.

Nation, P. (2000). Learning vocabulary in lexical sets: dangers and guidelines. *TESOL Journal*, 6 – 10.

Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical retrieval is not by competition: evidence from the blocked naming paradigm. *Journal of Memory and Language*, *76*, 253–272.

Nickels, L. (2001). Spoken word production. In B. Rapp (Ed.), *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind* (pp. 291–320). Philadelphia: Psychology Press.

Nozari, N., Freund, M., Breining, B. L., Rapp, B. C., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*. doi:10.1080/23273798.2016.1157194

O'Séaghdha, P. G., & Frazer, A. K. (2014). The exception does not rule: attention constrains form preparation in word production. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *40*, 797–810.

O'Séaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming : sequential and nonsequential processes in word production. *Journal of Experimental Psychology. Human Perception and Performance*, *26*, 57–73.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*, 227–52.

Orjada, S., & Beeson, P. M. (2005). Concurrent treatment for reading and spelling in aphasia. *Aphasiology*, *19*, 341–351.

Page, M., Wilson, B. a., Shiel, A., Carter, G., & Norris, D. (2006). What is the locus of the errorless-learning advantage? *Neuropsychologia*, *44*, 90–100.

Papathanasiou, E. (2009). An investigation of two ways of presenting vocabulary. *ELT Journal*, *63*, 313–322.

Pashler, H., & Rohrer, D. (2007). Enhancing learning and retarding forgetting: choices and consequences. *Psychonomic Bulletin & …*, *14*, 187–193.

Perani, D., Cappa, S. F., Schnur, T. T., Tettamanti, M., Collina, S., Rosa, M. M., & Fazio, F. (1999). The neural correlates of verb and noun processing. a pet study. *Brain : A Journal of Neurology*, *122 (Pt 1*, 2337–44.

Phillips, J. a, Noppeney, U., Humphreys, G. W., & Price, C. J. (2002). Can segregation within the semantic system account for category-specific deficits? *Brain : A Journal of Neurology*, *125*, 2067–80.

Pilgrim, L. K., Fadili, J., Fletcher, P. C., & Tyler, L. K. (2002). Overcoming confounds of stimulus blocking: an event-related fmri design of semantic processing. *NeuroImage*, *16*, 713–723.

Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 529–536.

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: behavioral and computational studies. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 393–406.

Rapp, B. C. (2002). Uncovering the cognitive architecture of spelling. In A. E. Hillis (Ed.), *The Handbook of Adult Language Disorders* (pp. 47–69). New York: Psychology Press.

Rapp, B. C., Benzing, L., & Caramazza, A. (1997). The autonomy of lexical orthography. *Cognitive Neuropsychology*, *14*, 71–104.

Rapp, B. C., & Caramazza, A. (1997). The modality-specific organization of grammatical categories: evidence from impaired spoken and written sentence production. *Brain and Language*, *286*, 248–286.

Rapp, B. C., & Fischer-Baum, S. (2014). Representation of orthographic knowledge. In V. Ferreria, M. Goldrick, & M. Miozzo (Eds.), *Oxford Handbook of Language Production*. New York: Oxford University Press.

Rapp, B. C., Fischer-Baum, S., & Miozzo, M. (2015). Modality and morphology: what we write may not be what we say. *Psychological Science*, *26*, 892–902.

Rapp, B. C., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*, 460–499.

Rapp, B. C., & Gotsch, D. (2001). Spelling disorders: cognitive theory in clinical practice. In R. S. Berndt (Ed.), *Handbok of Neuropsychology* (2nd ed., pp. 221–235). Amsterdam: Elsevier Science.

Rapp, B. C., & Kane, A. (2002). Remediation of deficits affecting different components of the spelling process. *Aphasiology*, *16*, 439–454.

Rapp, B. C., & Kong, D. (2002). Revealing the component functions of the graphemic buffer. *Brain and Language*, *83*, 112–114.

Rapp, B. C., & Lipka, K. (2011). The literate brain: the relationship between spelling and reading. *Journal of Cognitive Neuroscience*, *23*, 1180–1197.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in english: are they real? do they matter? *Cognitive Psychology*, *53*, 97–145.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: the arc nonword database. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *55*, 1339–62.

Raymer, A., Cudworth, C., & Haley, M. (2003). Spelling treatment for an individual with dysgraphia : analysis of generalisation to untrained words. *Aphasiology*, *17*, 607–624.

Raymer, A., Strobel, J., Prokup, T., Thomason, B., & Reff, K.-L. (2010). Errorless versus errorful training of spelling in individuals with acquired dysgraphia. *Neuropsychological Rehabilitation*, *20*, 1–15.

Rips, L., Shoben, E. J., & Smith, E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory . *Perspectives on Psychological Science*, *1*, 181–210.

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255.

Roelofs, A. (1999). Phonological segments and features as planning. *Language and Cognitive Processes*, *14*, 173–200.

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 922–939.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481–498.

Roux, S., & Bonin, P. (2009). Neighborhood effects in spelling in adults. *Psychonomic Bulletin & Review*, *16*, 369–373.

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58.

Sage, K., & Ellis, A. W. (2006). Using orthographic neighbours to treat a case of graphemic buffer disorder. *Aphasiology*, *20*, 851–870.

Schneider, V. I., Healy, A. F., & Bourne, L. E. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention* (pp. 77–90). Mahwah, NJ: Lawrence Erlbaum Associates.

Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What is learned under difficult conditions is hard to forget: contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, *46*, 419–440.

Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory and Language*, *75*, 27–44.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: evidence from aphasia. *Journal of Memory and Language*, *54*, 199–227.

Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: convergent neuroimaging and neuropsychological evidence for the function of broca's area. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 322–7.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture-word interference studies. *Journal of Memory and Language*, *29*, 86–102.

Sevald, C., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, *53*, 91–127.

Shallice, T., & Warrington, E. K. (1984). Category specific semantic impairments. *Brain*, *107*, 829–853.

Sharps, M. J., Wilson-Leff, C. A., & Price, J. L. (1995). Relational and item-specific information as determinants of category superiority effects. *Journal of General Psychology*, *122*, 271–285.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology. Human Learning and Memory*, *5*, 179–187.

Shen, X. R., Damian, M. F., & Stadthagen-Gonzalez, H. (2013). Abstract graphemic representations support preparation of handwritten responses. *Journal of Memory and Language*, *68*, 69–84.

Squires, E. J., Hunkin, N. M., & Parkin, A. J. (1997). Errorless learning of novel associations in amnesia. *Neuropsychologia*, *35*, 1103–1111.

Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning spanish as a foreign language. *Bilingualism: Language and Cognition*, *15*, 490–502.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research : JSLHR*, *49*, 1175–1192.

Storkel, H. L., Bontempo, D. E., & Pak, N. S. (2014). Online learning from input versus offline memory evolution in adult word learning: effects of neighborhood density and phonologically related practice. *Journal of Speech, Language, and Hearing Research*, *57*, 1708–1721.

Sullivan, M. P., & Riffel, B. (1999). The nature of phonological encoding during spoken word retrieval. *Language and Cognitive Processes*, *14*, 15–45.

Tailby, R., & Haslam, C. (2003). An investigation of errorless learning in memory-impaired patients: improving the technique and clarifying theory. *Neuropsychologia*, *41*, 1230–1240.

Tainturier, M.-J., Moreaud, O., David, D., Leek, E. C., & Pellat, J. (2001). Superior written over spoken picture naming in a case of frontotemporal dementia. *Neurocase*, *7*, 89–96.

Tainturier, M.-J., & Rapp, B. C. (2001). The spelling process. In B. C. Rapp (Ed.), *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind* (pp. 263–290). Philadelphia: Psychology Press.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning : stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*, 45–58.

Terrace, H. S. (1963). Discrimination learning with and without “errors”. *Journal of the Experimental Analysis of Behavior*, *6*, 1–27.

Thiel, L., & Conroy, P. (2014). A comparison of errorless and errorful therapies for dysgraphia after stroke. *Neuropsychological Rehabilitation*, *24*, 172–201.

Tinkham, T. (1993). The effect of semantic clustering on the learing of second language vocabulary. *System*, *21*, 371–380.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*, 138–163.

Toglia, M. P., Hinman, P. J., Dayton, B. S., & Catalano, J. F. (1997). The blocked-random effect in pictures and words. *Perceptual and Motor Skills*, *84*, 976–8.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381–391.

Tuomiranta, L., Grönholm-Nyman, P., Kohen, F., Martin, N., & Laine, M. (2011). Learning and maintaining new vocabulary in aphasia: two controlled case studies. *Aphasiology*, *25*, 1030–1052.

Tuomiranta, L., Rautakoski, P., Rinne, J. O., Martin, N., & Laine, M. (2012). Long-term maintenance of novel vocabulary in persons with chronic aphasia. *Aphasiology*, *26*, 1053–1073.

Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology*, *5*, 275–287.

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences*, *5*, 244–252.

Underwood, B. J., Ekstrand, B. R., & Keppel, G. (1965). An analysis of intralist similarity in verbal learning with experiments on conceptual similarity. *Journal of Verbal Learning and Verbal Behavior*, *4*, 447–462.

Waring, R. (1997). The negative effects of learning words in semantic sets: a replication. *System*, *25*, 261–274.

Warrington, E. K., & McCarthy, R. a. (1983). Category specific access dysphasia. *Brain : A Journal of Neurology*, *106 (Pt 4*, 859–78.

Warrington, E. K., & McCarthy, R. a. (1994). Multiple meaning systems in the brain: a case for visual semantics. *Neuropsychologia*, *32*, 1465–73.

Wheeldon, L., & Monsell, S. (1994). Inhibition of spoken word production by priming a semantic competitor. *Journal of Memory and Language*, *33*, 332–356.

Wheeler, M. a, Ewers, M., & Buonanno, J. F. (2003). Different rates of forgetting following study versus test trials. *Memory (Hove, England)*, *11*, 571–580.

Zhang, Q., & Damian, M. F. (2010). Impact of phonology on the generation of handwritten responses: evidence from picture-word interference tasks. *Memory & Cognition*, *38*, 519–28.

CURRICULUM VITAE

Bonnie Breining was born on February 8, 1987 in Columbus, Ohio and spent much of her formative years in Austin, Texas. She received a Bachelor of Arts with honors in Cognitive Sciences and Psychology from Rice University in 2008, followed by a Master of Science with distinction in Cognitive Neuropsychology and Cognitive Neuroscience from the University of London in 2009 and a Master of Arts in Cognitive Science from Johns Hopkins University in 2011. Upon completion of her Ph.D., she will begin a postdoctoral appointment in the Department of Neurology at the Johns Hopkins University School of Medicine.